

Perceptual Rendering of Participating Media

VERONICA SUNDSTEDT

University of Bristol

DIEGO GUTIERREZ and OSCAR ANSON

University of Zaragoza

and

FRANCESCO BANTERLE, and ALAN CHALMERS

University of Bristol

High-fidelity image synthesis is the process of computing images that are perceptually indistinguishable from the real world they are attempting to portray. Such a level of fidelity requires that the physical processes of materials and the behavior of light are accurately simulated. Most computer graphics algorithms assume that light passes freely between surfaces within an environment. However, in many applications, we also need to take into account how the light interacts with media, such as dust, smoke, fog, etc., between the surfaces. The computational requirements for calculating the interaction of light with such participating media are substantial. This process can take many hours and rendering effort is often spent on computing parts of the scene that may not be perceived by the viewer. In this paper, we present a novel perceptual strategy for physically based rendering of participating media. By using a combination of a saliency map with our new extinction map (X map), we can significantly reduce rendering times for inhomogeneous media. The visual quality of the resulting images is validated using two objective difference metrics and a subjective psychophysical experiment. Although the average pixel errors of these metric are all less than 1%, the subjective validation indicates that the degradation in quality still is noticeable for certain scenes. We thus introduce and validate a novel light map (L map) that accounts for salient features caused by multiple light scattering around light sources.

Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism-color, shading, shadowing, and texture

General Terms: Algorithms, Performance, Experimentation, Human Factors

Additional Key Words and Phrases: Participating media, selective rendering, perception, saliency map, extinction map, light map, attention

ACM Reference Format:

Sundstedt V., Gutierrez, D., Anson, O., Banterle, F., and Chalmers, A. 2007. Perceptual rendering of participating media. *ACM Trans. Appl. Percpt.* 4, 3, Article 15 (November 2007), 22 pages. DOI = 10.1145/1278387.1278389 <http://doi.acm.org/10.1145/1278387.1278389>

This research was partly sponsored by the European Union within the CROSSMOD project (EU IST-014891-2) and the Spanish Ministry of Education and Research through the project TIN2004-07672-C03-03.

Authors' addresses: Veronica Sundstedt, Francesco Banterle, and Alan Chalmers, Department of Computer Science, Merchant Venturers Building, Woodland Road, BS8 1UB, Bristol, UK; email: veronica@cs.bris.ac.uk. Diego Gutierrez and Oscar Anson, Grupo de Informatica Grafica Avanzada (GIGA), Departamento de Informatica e Ingenieria de Sistemas, Universidad de Zaragoza, Edificio Ada Byron, Maria de Luna 1, 50018, Zaragoza, Spain; email: diegog@unizar.es.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2007 ACM 1544-3558/2007/11-ART15 \$5.00 DOI 10.1145/1278387.1278389 <http://doi.acm.org/10.1145/1278387.1278389>

ACM Transactions on Applied Perception, Vol. 4, No. 3, Article 15, Publication date: November 2007.

1. INTRODUCTION

Rendering of physically based imagery using global illumination techniques is very useful and necessary in areas, such as safety, military, or industry [Rushmeier 1994]. The applications in these areas usually deal with the analysis of visual perception under unfavorable environmental conditions, where the presence of a medium have a noteworthy influence in the visibility (and, therefore, the design) of certain elements, such as road signs, fire exit signs, car headlamps, and road lighting. Examples of these participating media include smoke, fog, dust, flames, silty and abyssal waters or atmospheric phenomena, as shown in Figure 1. Simulating light interaction with this participating media implies the correct computation of the absorption and scattering of the light and is thus computationally very expensive.

In traditional global illumination rendering, not all computations performed effect how the final image is perceived. To improve the efficiency of rendering it is necessary to perform less rendering work, while not affecting the perception of the resultant image. Perceptually based rendering algorithms reduce the computational cost by exploiting limitations of the human visual system. For example, the amount of information that reaches our eyes is far greater than our brain is able to process. Selective visual attention is a complex action, which is used to find and focus on relevant information in a quick and efficient manner. Perceptually based metrics can be used as a stopping criterion for rendering by determining if continuing the computation would lead to a perceptual improvement. Selective rendering algorithms generate images where the rendering quality is varied spatially across the image [Sundstedt et al. 2005b; Chalmers et al. 2006]. Areas of high perceptual importance will have a more accurate solution. However, the difference in image quality will not be perceived by the viewer.

To significantly reduce physically based rendering times for participating media, we propose a novel perceptual strategy based on the combination of a saliency map [Itti et al. 1998] with our new extinction map (X map) and light map (L map). The X map stores in image-space the exponential decay of light in the medium. The L map accounts for salient features caused by multiple light scattering around light sources. This combination, that we have called the XSL map, is used to guide a selective rendering of a scene, with more accurate estimates in the perceptually important areas. The novelties of our work can be summarized as the introduction of the X-, and L-map concept and its combination with a modulated saliency map (S map) to guide a perception-based renderer for inhomogeneous participating media. In addition, two objective metrics are used to validate our approach and the perceived quality of selective rendering results is assessed by two forced-choice preference experiments.

The rest of the paper is organized as follows: In Section 2, we present a short summary of how visual perception has been incorporated in perceptually-based rendering algorithms. In Section 3, the physics of participating media that underpins this work is presented. We describe our novel selective rendering system using the XSL map in Section 4, and present the results in Section 5. We validate the XSL map in Section 6 using both objective metrics and subjective experiments. In Section 7, we conclude the work presented in this paper and discuss ideas for future research.

2. RELATED WORK

Early work in perceptually assisted rendering was mainly concerned with the acceleration of ray tracing. Although techniques based on visual attention had been developed before, such as adaptive sampling [Mitchell 1987] for ray tracing, which applied perceptual considerations, have recently become more popular. An good overview of early perceptually driven radiosity methods is given in by Prikryl [2001]. A frequency-based ray tracer using a more complete model of the human visual system, which incorporated the visual system's spatial processing behavior and sensitivity change as a function of luminance was developed by Bolin and Meyer [1995]. An even more sophisticated model incorporating visual



Fig. 1. Real photograph of a road with fog showing the effect of light scattering.

masking was developed by Ferwerda et al. [1997]. Visual difference predictors have been used both to direct the next set of samples within stochastic ray tracing, and as a stopping criteria [Myszkowski 1998; Bolin and Meyer 1998]. These algorithms both required repeated applications of the perceptual error metric, which was an expensive operation. The cost of such metrics were reduced in Ramasubramanian et al. [1999] by precomputing the spatial frequency component from a cheaper estimate of the scene image. Spatiotemporal sensitivity of the human visual system was later added to Daly's VDP to create a perceptually based animation-quality metric (AQM) [Myszkowski 2002], which was used to guide a hybrid ray tracing and image-based rendering (IBR) approach to improve rendering performance in a key frame-based animation sequence. Myszkowski's framework assumed that the eye tracks all objects in a scene and the AQM added computational overhead to the rendering process.

Later, a visual attention model was used in Yee et al. [2001] to improve the efficiency of indirect lighting computations for dynamic environments. Yee et al. [2001] exploited a saliency model, termed the Aleph map, to adjust the search radius accuracy of the interpolation of irradiance cache values. In Dumont et al. [2003] a perceptual metric for interactive applications was presented in order to reduce the memory required for each diffuse texture while keeping the best image quality. Another perceptual metric was used in a component-based rendering framework to predict the relative importance of each component as a function of the materials visible from the desired viewpoint [Stokes et al. 2004]. Saliency maps and the notion of task objects were used in a real-time renderer to identify the most salient objects for which to render the glossy and specular components Haber et al. [2001]. In Cater et al. [2003] and Sundstedt et al. [2005b] both task maps and saliency map were used to vary a number of rays shot per pixel in a global illumination environment.

This paper is an extended version of previous work [Anson et al. 2006], where a new extinction map (termed the X-map) is combined with the saliency map of a scene to guide a perceptually based rendering approach of participating media. One of the main limitations in the previous algorithm was that it did not account for the scattering around light sources, which could result in noisy artifacts because of poor sampling. We address this and other issues in this paper, performing new perceptual validation tests to assess the visual quality of the new selectively rendered images.

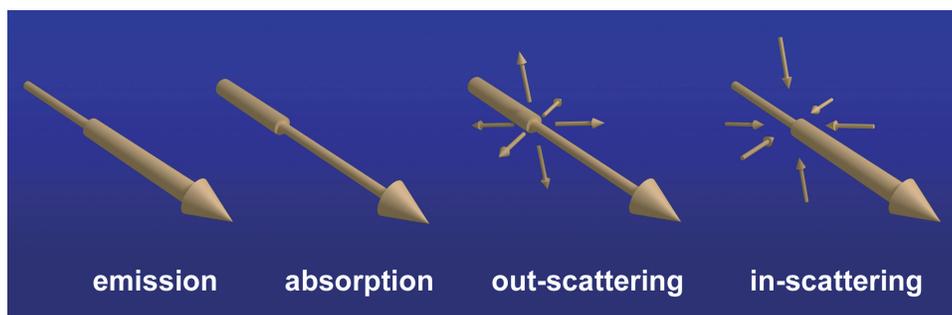


Fig. 2. The four new types of interaction of light in participating media (after Perez et al. [1997]).

3. LIGHT TRANSPORT IN PARTICIPATING MEDIA

Four new types of interactions may occur when light travels throughout a participating medium: emission, in-scattering, absorption, and out-scattering, as depicted graphically in Figure 2. The first two represent an increase of the light energy as it traverses the medium, whereas the last two represent a decrease. All these terms must be calculated ideally at each differential step, thus making it a very expensive process. This is the reason why participating media computations usually are ignored in light-transport algorithms, considering that light only interacts with the geometry present in the scene.

This approach is sufficient for the majority of situations, when light-scattering effects can be neglected from a perceptual perspective. The rendering equation introduced in computer graphics by Kajiya [1986] yields good solutions for those cases. However, there are situations where a complete modeling of the light transport is necessary (such as driving simulators in foggy environments or underwater simulations).

A physically accurate light-transport solution in such media implies solving the radiative transfer equation (RTE) [Glassner 1994], which is an integro-differential equation and noticeably more complex. This equation does include the four new types of interactions introduced previously, each of them governed by a particular wavelength-dependent coefficient. The equation describes the variation of radiance L_λ in a point x in the direction ω . Following the notation introduced in Jensen and Christensen [1998], it can be written as

$$\frac{\partial L_\lambda(x, \vec{\omega})}{\partial x} = \alpha_\lambda(x)L_{e,\lambda}(x, \vec{\omega}) + \sigma_\lambda(x)L_{i,\lambda}(x, \vec{\omega}) - \alpha_\lambda(x)L_\lambda(x, \vec{\omega}) - \sigma_\lambda(x)L_\lambda(x, \vec{\omega}) \quad (1)$$

where α_λ and σ_λ are the absorption and scattering coefficients, and $L_{e,\lambda}$ and $L_{i,\lambda}$ are the emitted and in-scattered radiance, respectively. To simplify the notation, the wavelength-dependency of the parameters is assumed and is omitted in the rest of the paper. For general, inhomogeneous media, both the absorption and scattering coefficients vary throughout the medium.

Defining the extinction coefficient as $\kappa(x) = \alpha(x) + \sigma(x)$ and knowing that the in-scattered radiance L_i depends on the incoming radiance from all possible directions $\vec{\omega}'$ into the point x over the sphere Ω mean, we can rewrite Eq. (1) in the form

$$\frac{\partial L(x, \vec{\omega})}{\partial x} = \alpha L_e(x, \vec{\omega}) + \sigma \int_{\Omega} p(x, \vec{\omega}', \vec{\omega}) L(x, \vec{\omega}') d\vec{\omega}' - \kappa L(x, \vec{\omega}) \quad (2)$$

where $p(x, \vec{\omega}', \vec{\omega})$ is the phase function that describes the fraction of radiance arriving from direction $\vec{\omega}'$ that is in-scattered along the path. Solving the integro-differential Eq. (2) by integrating with respect to its boundary conditions, we have the integral form of the RTE [Siegel and Howell 1992] given by

$$\begin{aligned} L(x, \vec{\omega}) &= e^{-\tau(x_0, x)} L(x_0, \vec{\omega}) \\ &+ \int_{x_0}^x e^{-\tau(x', x)} \alpha(x') L_e(x', \vec{\omega}) dx' \\ &+ \int_{x_0}^x e^{-\tau(x', x)} \sigma(x') \int_{\Omega} p(x', \vec{\omega}', \vec{\omega}) L(x', \vec{\omega}') d\vec{\omega}' dx' \end{aligned} \quad (3)$$

where $e^{-\tau(x', x)}$ is the transmittance along the optical path length $\tau(x', x)$, which express the attenuation of the light along its way. Optical length is defined as

$$\tau(x', x) = \int_{x'}^x \kappa(s) ds \quad (4)$$

As it can be seen in Eq. (3), the computation of the radiance at a point in the scene directly depends on the radiance of all other points in the medium surrounding it. This fact makes the light transport computation in participating media extremely complex and much more expensive than the traditional radiance equation. In the simplest case, where the medium is homogeneous, Eq. (3) could be simplified, since the absorption and scattering coefficients would not be functions of x' . Our work does not rely on this simplification.

We have so far considered only elastic scattering events, where photon interaction does not imply a transfer of energy (change of wavelength) between photons. To also compute inelastic scattering, where such energy transfers do occur, we need to develop the RTE equation further, by adding an inelastic scattering term, which can be expressed as a double integral over the domains of the solid angle and wavelength

$$\int_{\Omega} \int_{\lambda} \alpha_{\lambda_i}(x) f(x, \lambda_i \rightarrow \lambda) L_{\lambda_i}(x, \vec{\omega}') \frac{p_{\lambda}(x, \vec{\omega}', w)}{4\pi} d\vec{\omega}' d\lambda_i \quad (5)$$

where α_{λ_i} is the absorption coefficient for wavelength λ_i , $f(x, \lambda_i \rightarrow \lambda)$ is the function that governs the efficiency of the energy transfer between wavelengths, defined as the probability of a photon of λ_i being re-emitted at λ . Adding this term to the RTE (Eq. 3) we obtain the full radiative transfer equation (FRTE) [Gutierrez et al. 2005]. Since our proposed method to reduce rendering times is independent of the algorithm used to compute light transport in participating media, there is no restriction as to what type of scattering events can be taken into account.

4. PERCEPTUAL STRATEGIES

Given the complexity of the integro-differential equation that governs light transport in participating media, the time needed to render a physically accurate image is usually measured in hours. For complex scenes, the process can take as much as tens of hours. There is a lot of work done previously that deals with rendering of participating media at interactive rates [Sun et al. 2005]. Although, these methods are either based on clever hacks with no physical ground or they only consider single scattering computations instead of multiple scattering events.

Our goal is to compute scenes with physically accurate participating media in a reasonable time, while maintaining a high-fidelity perceptual result. To achieve this goal, we first build a series of maps of the scene under the desired viewpoint. This process only impose an overhead of less than a

few seconds per frame, but later helps us to reduce the rendering times by up to an order of magnitude. Each map contains different information about the scene to be rendered coded in grey scale: (1) the light extinction in the medium, (2) the saliency features, and (3) the scattering around light sources.

We term these maps the X, S, and L map, respectively. These three maps are then combined into one, which, as a result, yields information on what features of the scene are predicted to be visible, as a function of the three considered factors. The combined map is then used as a director to guide the renderer. In this way, computing efforts are concentrated only in the areas of the image, which are perceptually more important. The X, S, and L map and their combinations are described in the following sections.

4.1 The Extinction Map

In the general case of inhomogeneous participating media, the intensity of light as it transverses the medium is reduced by the transmittance $e^{-\tau(x',x)}$ (see Eq. 4), which can be evaluated by ray marching. Therefore, independently on how much irradiance reaches a given object, only a fraction of it will finally reach the viewer. The first idea of our work is to precompute that fraction beforehand, storing in image-space attenuation values in an extinction map, which we term the X map.

To obtain the X map, rays are cast from the eye into the scene, ray marching through the medium until the first intersection, and saving the result of the exponential decay in the X map. This represents the percentage of light (both direct and diffuse) that will reach the eye for each pixel (or subpixel, if the resolution of the map is increased). For homogeneous media, the attenuation is simply e^{-ks} and faster ray tracing can be used instead of ray marching. In either case, the distance to the intersection is also saved in a Z buffer. This allows for instant recomputations of the X map if the description of the medium in the scene changes (such as dynamic fog or smoke), since intersections do not need to be recalculated at each step. The bottom left image in Figure 3 shows the X map for the Lucy scene.

For outdoor scenes, we found that large areas with infinite depth (i.e., the sky) could occur. In this case, we can not calculate the X map, which is set to a value of zero. However, a medium can exist taking up all the space between that apparent empty background and the camera. In this way, we can have a very poor-quality render of such medium that can be noticed by viewers. To overcome this problem, a technique is introduced, which we term *MinBound*. The first step in this method lies in calculating the minimum value x_{\min} in the X map. This value is then assigned to the zero values in the X map. This leads to an improvement that avoids rapid changes in quality between the ground and the sky in a scene.

4.2 The Modulated Saliency Map

The saliency generator [Itti et al. 1998] is based on a bottom-up, feature-based model, which extracts features from an image that humans would automatically direct attention to. The saliency map is obtained from an estimate (snapshot) of the scene, where only direct lighting is computed using ray tracing [Longhurst et al. 2005; Sundstedt et al. 2005b], as shown in the top middle image in Figure 3. The saliency map is obtained via a combination of three conspicuity maps of intensity, color, and orientation. These conspicuity maps are computed using feature maps at varying spatial scales. The features can be thought of as stimuli at varying scales and conspicuity as a summary of a specific stimulus at all the scale levels combined. Saliency, on the other hand, can be seen as a summary of all the conspicuity of all the stimuli combined together. For further details on the computational model of the saliency map, we refer the reader to the work by Itti et al. [1998]. This process takes only a few seconds, although a hardware implementation can generate a saliency map in the order of tens of milliseconds [Longhurst et al. 2006].

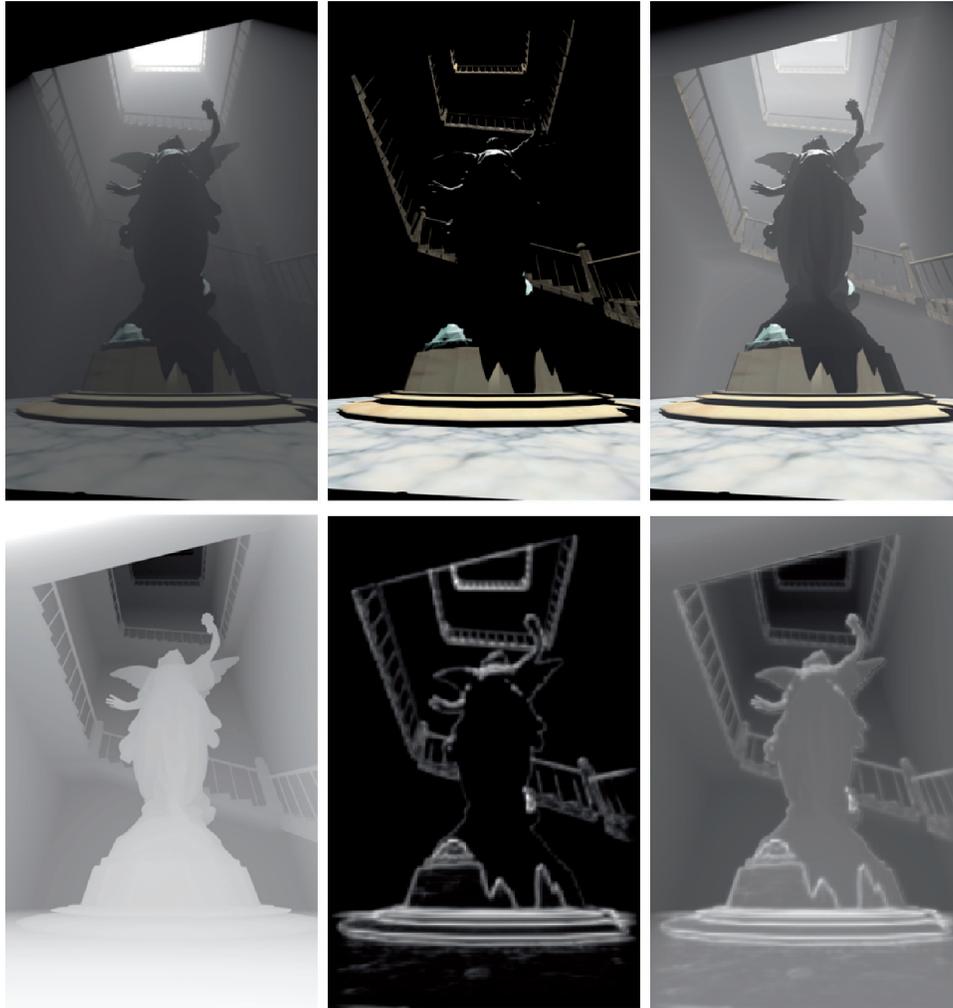


Fig. 3. The top Lucy scene: (left) gold standard, (middle) initial fast ray-traced snapshot (direct lighting and no medium), and (right) a false estimate of the medium, computed from the combination of the ray-traced snapshot and the inverted X map. The bottom Lucy scene: (left) the X map, (middle) the S map, and (right) the map.

However, in the presence of a participating media, salient features are modulated by the scattering and emission of light. This is taken into account in our technique by modulating the saliency map accordingly. We term this modulated saliency map—S map. The bottom middle image in Figure 3 shows the S map for the Lucy scene. We found that extinction and visibility can roughly be considered inversely proportional for our purposes. The X map is thus first inverted and then added to the snapshot of the scene, applying the saliency algorithm to the resulting combination. The result of this process for the Lucy scene is shown in the top right image in Figure 3. The X map, is, therefore, used for two different purposes. First, we use it to estimate the extinction in the scene because of the medium, and second, it is inverted and combined with the snapshot to act as input to the saliency map generator. While this is a not a physically based accurate method, we have empirically determined that the invert

X map operation suits our needs in the perceptually-based selective rendering system. The validation tests described in Section 6 confirm this.

4.3 The XS Map

Once the X and the S maps are obtained, they are combined into what we term the XS map. The XS map intuitively yields information on the salient features of the scene modulated by the extinction in the participating medium. For the combination, we follow a similar approach as presented in Sundstedt et al. [2005b]. Our combination is given by $XS(w_x, w_s, op)$, where XS represents the combined S and X maps, and w_x and w_s are coefficients applied to the values in the respective maps, allowing for different weighting to define relative importance of the maps. The coefficients are combined through the operator op . Selection of appropriate operator op controls the combination of the two maps in the selective renderer. We used an add operator to combine the information from both maps such that all weighted features are preserved. An equal weighting of both maps would, therefore, be $XS(0.5, 0.5, +)$, which correspond to our XS map, rendered images, $XS(1, 0, +)$ to the images rendered only with the X map, and $XS(0, 1, +)$ to the images rendered only with the saliency map. A multiplicative operator could also be used in order to represent saliency modulated by the decay of light in the media. In any case, the XS map will guide the selective rendering process by spending computational resources in areas of higher perceptual importance and low extinction. The bottom-right image in Figure 3 shows the XS map for the Lucy scene.

4.4 The Light Map

The XS map seems to be sufficient if there are not visible light sources against an empty background. In the presence of a participating medium, multiple light scattering around light sources becomes an important salient feature and this will slip by undetected in the case described. This can be seen in Figure 1, where areas of strong scattering around light sources become a salient feature of the whole scene. In the first experiment presented in Section 6, we also show, using a visualization of eye-gaze data, that the participants did attend to areas with light scattering.

However, neither the X map nor the S map (and, therefore, neither its combination into the XS map) will identify these areas as visually important. This will lead to poor sampling during the rendering stage, resulting in clearly visible artifacts in the image. The artifacts can be seen in the bottom right image in Figure 4 rendered using the XS map. The X, S, and XS map are shown on the top and neither of these maps predict the scattering around the light sources, which become a key feature of the scene. The bottom left image shows the gold standard for the road scene. Even though the image on the right took an order of magnitude less to render than the gold standard, artifacts around light sources caused by poor sampling are clearly visible. This is because of the maps wrongly indicating the rendering algorithm to undersample these areas.

To overcome this problem we propose a novel solution by introducing a new map, which we term the light map (L map). This map is obtained by first identifying the light sources in the scene and then producing a scan-line image in which light sources appear constant white, while the rest of the scene renders black. This render is termed the *0-LightMap* (L_0) and an example for the road scene is shown in Figure 5a. We then want to define an area around those light sources where the scattering of light will become an important feature of the scene. This problem is similar to applying a point spread function to the white pixels rendered in the 0-LightMap. Since, at this stage, we only want to identify future salient areas (not render the actual multiple scattering), we can assume that the shape of this point spread function depends only on the optical thickness of the media. More accurate shapes could be considered by following the work of Narasimhan and Nayar [2003], for example. Although, our interest is to keep the method as computationally simple as possible, while still producing good results.



Fig. 4. The top road scene: (left) the X map, (middle) the S map, and (right) the XS map. The bottom road scene: (left) the gold standard and (right) the image rendered when applying the XS map. Notice how the scattering around light sources is poorly sampled, resulting in visible artifacts.

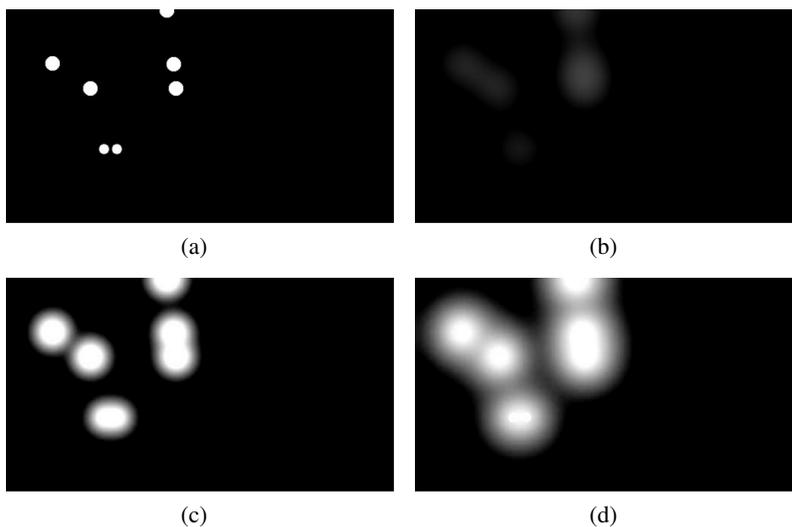


Fig. 5. Light maps for the road scene: (a) the 0 -LightMap (L_0), (b) a filtered L_0 using a Gaussian filter with $\sigma = 24$, (c) an example of our proposed multi-Gaussian filter with $n = 8$ and $\sigma_{\text{base}} = 4$, and (d) an example of our proposed multi-Gaussian filter with $n = 12$ and $\sigma_{\text{base}} = 6$.

As an initial candidate, we could chose a Gaussian function, with a wider kernel as the medium thicken. However, this would result in low values around the light sources, as can be seen in Figure 5b. The 0 -LightMap can be computed in a high dynamic-range (HDR) format to ensure that its convolution with a Gaussian filter yields high enough values. In order to make our method more general, we chose

to use a multi-Gaussian filter, as can be seen in Figure 5c and d:

$$L_n = \begin{cases} L_0 & i = 0 \\ L_{i-1} + (L_{i-1} \star G(\sigma_{\text{base}}i)) & i > 0 \end{cases} \quad (6)$$

where L_n is the n -*LightMap*, G is a Gaussian function with standard variance $\sigma_{\text{base}}i$, which is convolved with the result of the previous step L_{i-1} . The values n and σ_{base} are user-defined parameters, directly related to the thickness of the medium. Intuitively, a very thick medium will need a higher value for n and a medium-high value for σ_{base} , whereas lower values will do for both parameters as the medium gets thinner. We found that values of n between 4–8 work well for moderately thin media, increasing as high as 20 for thicker media. For σ_{base} , values of up to 4 work well in thin media, increasing as high as 10 as the medium thickens. Figure 5 summarizes this for the road scene, where the initial 0-LightMap is obtained first. The result of convolving it with a Gaussian kernel is shown next, where it can be seen how the convolution yields low values. Finally, two different maps are shown, using the multi-Gaussian approach with a thin medium with $n = 8$ and $\sigma_{\text{base}} = 4$ and a thicker medium with $n = 12$ and $\sigma_{\text{base}} = 6$.

4.5 The XSL Map

In order to generate the final director map, which we term the XSL map, the previous computed XS map is combined with the L map. The XSL map gives modulated information about the salient features of the scene and the extinction and scattering produced by the present participating media. For the combination we follow the same technique as for the XS map, combining the XS map and the L map through the term $XSL(w_{xs}, w_l, op)$, where XSL represents the combined XS map and L map, and w_{xs} and w_l the coefficients applied to the values in the XS map and the light map, respectively, which allow for different weighting of the maps. The coefficients are combined through the operator op . Selection of appropriate operator op controls the combination of the two maps in the selective renderer. An add operator was also used in the combination of both maps such that all weighted features are preserved. In order to give the same weight to both maps we use the combination term as $XSL(0.5, 0.5, +)$, which correspond to our XSL map, rendered images. If we want to render images only with the XS map, we will use $XSL(1, 0, +)$ and $XSL(0, 1, +)$, if on the contrary, we use only the L map. The XSL-map then work as a director map in the rendering process, favoring the quality where the perceptual saliency is high, the extinction is low, and where the scattering around the light sources is strong. The bottom-right image in Figure 6 shows the XSL map for the road scene.

5. RESULTS

The presented perceptual strategies have been used to extend our own global illumination rendering engine. Although, our algorithm for generating the XSL map can be incorporated into any rendering system, which then is used to direct the rendering parameters. In our engine the maximum number of rays per pixel is input as a user-defined parameter. For each pixel to be shot, this number is weighted by the corresponding value in the XSL map. Higher values in the director map will result in more rays shot per pixel and as a result of this an increased quality with lower aliasing effects. The system could easily be extended to support other parameters, such as the size step for ray marching the medium or the estimation quality in a volume photon map approach.

We rendered images from four different scenes to construct the experiment stimuli used in the perceptual validations. These scenes are termed balls A, balls B, Lucy, and road. For the first experiment presented in Section 6, we rendered four versions for each scene: (1) a gold standard, (2) using only the X map, $XS(1, 0, +)$, (3) using only the saliency map, $XS(0, 1, +)$, and (4) using the XS map. In the gold standard image, sixteen rays were shot for each pixel in the whole image following the approach

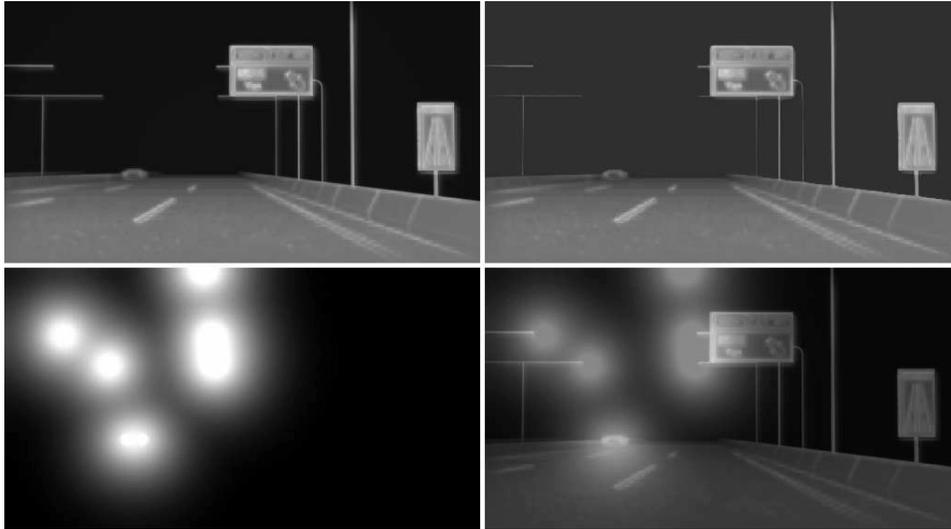


Fig. 6. The road scene: (top left) the normal XS map, (top right) the XS map using the *MinBound* method, and (bottom left) the L map, and (bottom right) the XSL map.

by Sundstedt et al. [2005a]. We also refer to this as our high-quality reference solution, or gold standard, which is defined as a hypothetical perfect rendering [Woolley et al. 2003].

For the three remaining versions, a sixteen rays per pixel ceiling was established. For the second experiment presented in Section 6, we also rendered an additional image for the Road scene using our new XSL map with the *MinBound* method. This was done because of the artifacts caused by the light scattering in this particular scene. The experiment stimuli created using the new XSL map can be seen in Figure 7.

All images were rendered on a PC with a 2, 6 GHz Pentium 4 processor, based on Windows XP, and at 512 pixels in their largest side. Figure 8 shows the timing comparison (in hours) between the reference gold standard and the solution generated using the XS maps. Rendering the gold standard images takes between 5.2 and 10.3 times longer than using our XS map (and the XSL map for the road scene). Anson et al. [2006] showed that even though the XS map combination takes approximately 1.15 longer to render than using only the S map, the visual quality that is obtained is slightly better, allowing us to afford that extra rendering time. On the other hand, rendering the road scene with the XS map is approximately twice as fast than using our new XSL map with the *MinBound* method. Using the XSL map is slower (16.21 versus 31.15 hours), although the visual quality of the rendered scene is clearly better than the quality obtained by using the XS map. Figure 7 shows a comparison between the road scene images rendered using the XS map and the XSL map.

6. PERCEPTUAL VALIDATION

The perceptual validation consists of two experiments. The first experiment was based upon four scenes, namely: balls A, balls B, Lucy, and road. Examples of the stimuli can be seen in Figure 9. Two images were rendered for each of the four scenes. One was rendered in high quality (gold standard) and the other one was rendered selectively using the XS map. The second experiment evaluates the visual quality of the selectively rendered stimuli generated using our new XSL map for the road scene. Two new images were rendered for this scene using the techniques described in Section 5. The same gold standard of the road scene was used, as in the first experiment.



Fig. 7. Road scene rendering comparisons: (top) XSL map rendered image, (bottom left) zoomed in XS map rendered image, (middle) gold standard, and (bottom right) zoomed in XSL map rendered image.

Two techniques were used to evaluate the visual quality of the selectively rendered stimuli. The first technique used two objective difference metrics to find the percentage of pixels, which are in error (and the average pixel error) between the gold standard stimuli and the stimuli selectively rendered using the XS or XSL map. The second technique was undertaking subjective psychophysical validations. In these validations, human subjects performed a two-alternative forced-choice (2AFC) preference experiment. Our hypothesis in the experiments was that the participants would not be able to discern a difference between the selectively rendered stimuli and the gold standard.

6.1 Objective Difference Metrics

The first metric used for comparing the differences were mean square error (MSE). A difference map was generated using MSE, which then was averaged for all pixels. The second metric used was Daly's perceptually based visual differences predictor (VDP) [Daly 1993]. The VDP takes two images as input and applies a model of human vision so that each image is represented visually. The VDP then generates a difference map (per pixel) of the probability (P) that, at any given pixel in the image, a human observer will perceive a difference between the two images.

We report the percentage of pixels for which the probability of detection between images is over 0.75. Myszkowski [1998] report that this is a standard threshold value for discrimination tasks. The difference map was then averaged to obtain the error value for each image pair. Although the average pixel error refers to a value, which is comparable between the different cases, it is not directly related to the error-detection probability [Longhurst et al. 2005]. For example, the same average could refer to

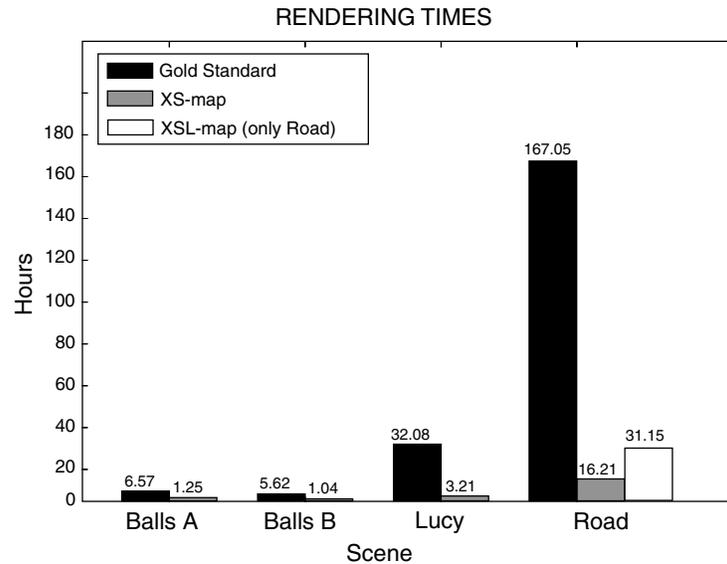


Fig. 8. Timing comparison for the four scenes between the reference gold standard solution and the image generated using the XS map (and the XSL map for the road scene). Timings include the preprocessing step of the XS map and XSL map generation.

an imperceivable error throughout the image or a highly visible error in a specific location [Longhurst et al. 2005].

The MSE and VDP average percentage pixel error results can be seen in Table I. Both average pixel error values were less than 1% for each of the four scenes using the XS map. Another way to display the results are the percentage of pixel errors for each scene, as shown in Table II. The percentage of error pixels using VDP was 0.57% or less using the XS map. The percentage of MSE error pixels for the balls A and balls B scenes were around 3%. For the Lucy and road scenes these increased to 8.5 and 50%, respectively. The MSE and VDP difference maps for the road scene using the XS map are shown in Figure 10.

The error levels were also computed between the gold standard and the XSL map rendered image for the road scene. The percentage of MSE error pixels was 45.16% and the average MSE error was 1.74%. The percentage of VDP error pixels for the same condition was 0.00% and the average pixel error 0.09%. These results show that the percentages of MSE and VDP error pixels are smaller compared to the errors in the images rendered using the XS map, as shown in Table II. Introducing the L map to compute the resulting XSL map can lead to an increase in the average pixel error on a per scene basis, as shown in Table I. This is because of the fact that more samples were computed near the light sources whereas other areas of the image were subsampled as a result. The MSE and VDP difference maps for the road scene using the XSL map are shown in the bottom left and right image of Figure 11.

6.2 Subjective 2AFC Experiments

Although objective visual quality metrics have been successfully used in the past to assess the pixel error of selective-rendered stimuli [Longhurst et al. 2005; Debattista and Chalmers 2005], it is important to validate the resulting images using subjective responses from human observers. Previous work has shown that VDP responses could replace human judgments in evaluating progressive radiosity solution convergence [Martens and Myszkowski 1998]. Even though the average pixel errors for all scenes were

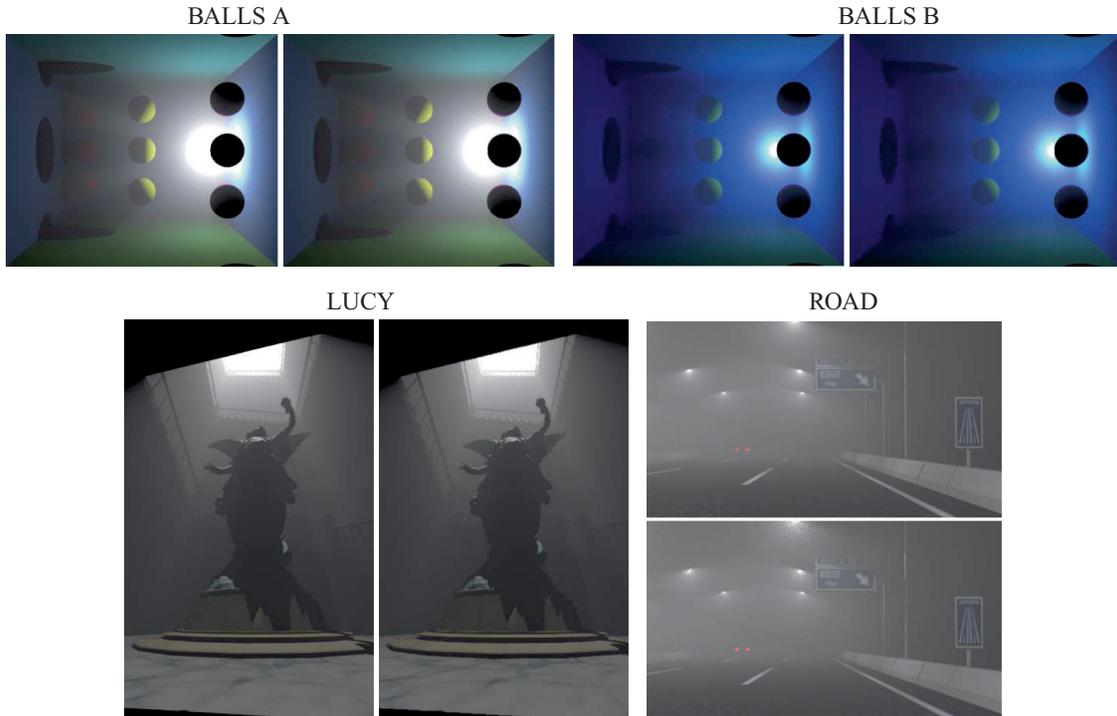


Fig. 9. Experiment stimuli in the first experiment: (top left) balls A, (top right) balls B, (bottom left) Lucy, and (bottom right) road. On the left side, the gold standard images are shown. On the right side, the XS map images are shown. In the case of the road scene, the gold standard is on top.

Table I. Average MSE and VDP Pixel Error between the Four Scenes Gold Standard and XS Map Renderings^a

	Balls A	Balls B	Lucy	Road	Road XSL
MSE	0.13%	0.10%	0.45%	0.87%	1.74%
VDP	0.27%	0.61%	0.76%	0.19%	0.09%

^aThe average pixel errors between the gold standard and the XSL map renderings are also shown in road XSL.

Table II. Percentage of MSE and VDP Error Pixels between the Four Scenes Gold Standard and the XS Map Renderings^a

	Balls A	Balls B	Lucy	Road	Road XSL
MSE	3.23%	3.04%	8.51%	50.33%	45.16%
VDP	0.35%	0.29%	0.57%	0.04%	0.00%

^aThe percentages of error pixels between the gold standard and the XSL map renderings are also shown in road XSL.

less than 1% using the XS map, a psychophysical experiment was run with 48 human observers. In this experiment participants performed a 2AFC, which assessed the perceived quality of selective rendering using the XS map. Within this experiment, two different display conditions for the experimental images were used, described in Section 6.2.1 and 6.2.2:

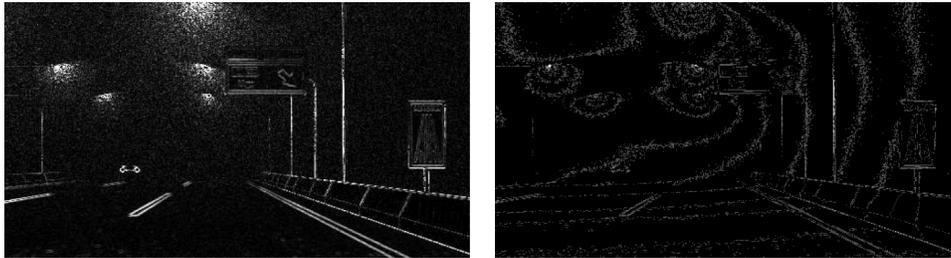


Fig. 10. Objective difference metrics: (left) mean square error (MSE) and (right) visual difference predictor (VDP). Comparison between gold standard and XS map image for the road scene. Brighter pixels in the image show greater perceptual differences. The bright concentric isocurves in the VDP image are produced as an effect of the constant ray-marching step used to render the medium.



Fig. 11. Experimental stimuli of the road scene used in the second experiment: (top left) the gold standard and (top right) the XSL map rendered image. Objective difference metrics: (bottom left) mean square error (MSE) and (bottom right) visual difference predictor (VDP). Comparison between gold standard and XSL map image for the road scene. Brighter pixels in the image show greater perceptual differences.

- *No reference*: discriminating which of two consecutively displayed images (gold standard and selectively rendered image in altered order) contained the worse rendering quality using a 2AFC task.
- *Reference*: discriminating which of two images (gold standard and selectively rendered image in altered order) were most similar to a gold standard. All three images in this condition were displayed at the same time, as shown in Figure 12.

In total, 48 participants (42 men and 6 women) in an age range of 19 to 38 took part in the first subjective experiment; 24 participants took part in each condition. In the second experiment, participants performed a 2AFC, which assessed the perceived quality of selective rendering using the XSL map. The second experiment only used the reference condition. This condition was chosen, based on the outcome of the first experiment. In total, 24 participants (19 men and 5 women) in an age range of 18 to 29 took part in the second visual trial. All subjects had a variety of experience with computer graphics and all

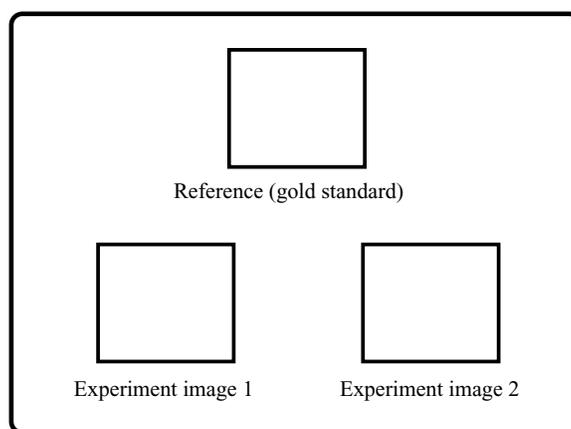


Fig. 12. 2AFC display used in the reference condition. At the top the reference image (gold standard) is shown. Below are the two experiment images (gold standard and selectively rendered image) randomized with one-half on each side.

self-reported normal or corrected-to-normal vision. All stimuli were presented on a 17" display monitor (1280 × 1024 resolution). The effects of ambient light were minimized and the lighting conditions were kept constant throughout the experiment. The participants were seated on an adjustable chair, with their eye level approximately level with the center of the screen, at a viewing distance of 60 cm. All stimuli were displayed on a screen, which had a 50% grey background.

6.2.1 No Reference Procedure. Before beginning the experiment, the subjects read a sheet of instructions on the procedure they were to perform. In the first condition, participants performed a 2AFC task between two consecutively displayed images. One of these images were the gold standard, whereas the other image was selectively rendered using the XS map. The consecutively shown images were displayed for 4 s each with a 50% grey image in between lasting for 2 s. After each trial, the participants were asked to judge which of the two images they thought contained the *worse* rendering quality. Participants circled their responses by choosing first or second on a questionnaire. One-half the high-quality images were on the left and one-half on the right, in random order. The first condition was chosen to study if a participant would be able to distinguish the image containing the worse rendering quality in the absence of a perceptually ideal solution. In reality, the XS map rendered image would not be displayed at the same time as a gold standard. To prevent repeated exposure to the same scene, each participant only performed one trial for each scene.

6.2.2 Reference Procedure. In the second condition, participants had to discriminate, using a 2AFC method [Neumann and Gegenfurtner 2002], which one of the two presented stimuli was *most similar* to a reference image (gold standard). The gold standard was always displayed on top in the center. The selectively rendered image and the gold standard were displayed below, side by side, as shown in Figure 12. The three images were displayed during 8 s. One-half of the gold standard images were on the left and one-half on the right, in random order. After each trial, participants circled their responses by choosing left or right on a questionnaire. The second condition was used to study if the outcome of having a gold standard for comparison would differ from the results in the no reference condition. Our hypothesis was that participants would be more likely to notice the differences in the presence of a gold standard. This is perhaps the fairest comparison to perform if one want to claim that images are perceptually indistinguishable from a fully converged solution. Although, one could also argue that this

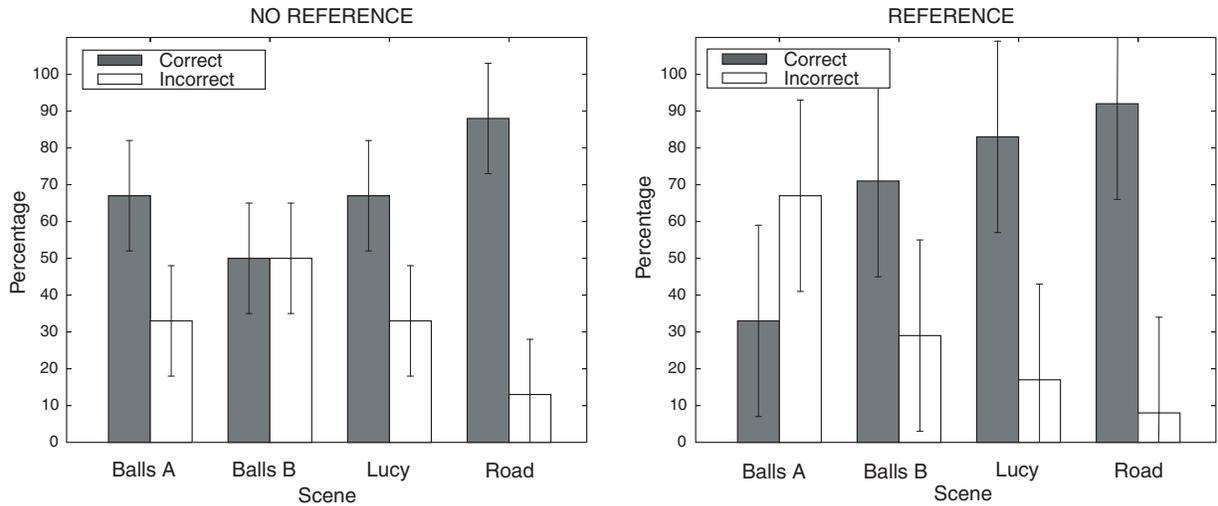


Fig. 13. Experiment results for the two conditions using the XS map: no reference versus reference with error bars representing standard deviation.

comparison is more likely to constitute a task in itself. This task could perhaps also alter the natural eye movements of human observers.

6.3 Results

Figure 13 shows the overall results of the first experiment. In each pair of conditions, a result of 50% correct selection in each case is the unbiased ideal. This is the statistically expected result in the absence of a preference or bias toward one case, and indicates that no differences between the gold standard and the selectively rendered images were perceived. For the first condition using the XS map, without a reference, the results show that 67% reported a correct result for the balls A and Lucy scene. For the balls B scene the percentage was 50% and in the road scene it was as high as 88%. When a reference image was introduced, the correct percentage for the balls A scene was 33%. For the balls B, Lucy, and road scenes the percentages increased to 71, 83, and 92%, respectively. Figure 14 shows the results of the second experiment. For the XSL map condition (using *MinBound*), the results show that 54% reported a correct result for the road scene.

6.4 Statistical Analysis and Discussion

The results were analyzed to determine any statistical significance. To find out whether the number of participants who correctly classified the worst visual quality image or the one most similar to a reference is what we would expect by chance, or if there was really a pattern of preference, we use a nonparametric technique called *Chi square*. A nonparametric test is used when the data is not normally distributed, as is the case for our binomially distributed data. A one-sample Chi square includes only one dimension, such as the case as in our experiments. The obtained (correct/incorrect) frequencies were compared to an expected 12/12 (24 for each condition) result to ascertain whether this difference would be significant. The Chi-square values were computed and then tested for significance, as shown in Table III. The obtained values show that when participants did not have a reference image, there was no significant difference between the gold standard and the XS map rendered images for the balls A, balls B, and Lucy scene ($p > 0.05$), as shown in Table III. This indicates that the participants not were able to correctly classify the image with worse rendering quality. For the road scene, the result

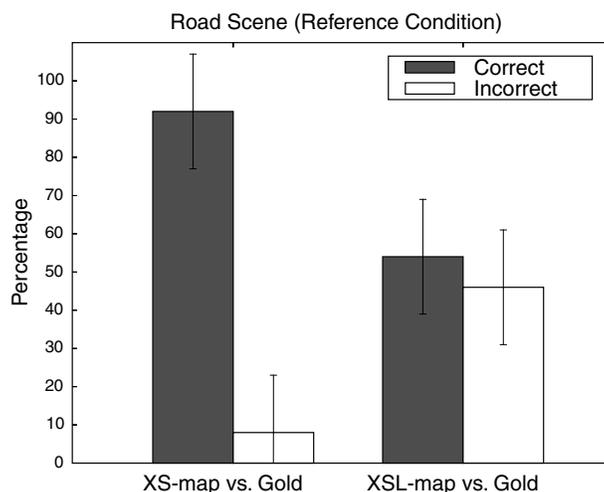


Fig. 14. Experiment results for the road scene while comparing the XS and XSL map to the gold standard. Error bars representing standard deviation.

Table III. Output for a Chi-Square Analysis

	No Reference	Reference
BALLS A	$\chi^2(1, N = 24) = 2.67, p = 0.10$	$\chi^2(1, N = 24) = 2.67, p = 0.10$
BALLS B	$\chi^2(1, N = 24) = 0, p = 1.00$	$\chi^2(1, N = 24) = 4.17, p = 0.04$
LUCY	$\chi^2(1, N = 24) = 2.67, p = 0.10$	$\chi^2(1, N = 24) = 10.67, p = 0.00$
ROAD	$\chi^2(1, N = 24) = 13.5, p = 0.00$	$\chi^2(1, N = 24) = 16.67, p = 0.00$
ROAD XSL		$\chi^2(1, N = 24) = 0.167, p = 0.683$

was highly significant ($p < 0.05$). In this scene, the participant could easily distinguish which of the two was the degraded one.

When the participants had a reference image to compare the experiment images with, the result was not significant only for the balls A scene ($p > 0.05$). This indicates that the participants were not able to correctly discriminate the gold standard from the two experimental images. For the balls B, Lucy, and road scene, the results were statistically different and not what would be expected by chance ($p < 0.05$). From these three scenes it can be concluded that the participants did manage to correctly choose the gold standard as most similar to the reference. These results from our first condition (no reference) are promising and show that the XS map can be used to produce rendered images with a high perceptual quality. The results from the second condition (reference) are also interesting since they indicate that high perceptual results can be obtained, but it is not necessarily true to claim that the XS map can produce images indistinguishable from a fully converged solution. Overall the results show that, for certain scenes, the participants managed to distinguish the worse quality or the image most similar to a reference. Both of these were true for the Road scene, in particular. The results presented in this section extend previous work [Longhurst et al. 2005] by showing that a low average pixel error of 1% not directly mean that an image will be indistinguishable from a gold standard. This shows the importance of using human observers in the validation process of realistic image synthesis algorithms.

In the second experiment, the difference in proportion between correct/incorrect answers was not significant, as shown in Table III. This indicates that the participants could not discern the difference between the gold standard and the selectively rendered stimuli created using the XSL map. By

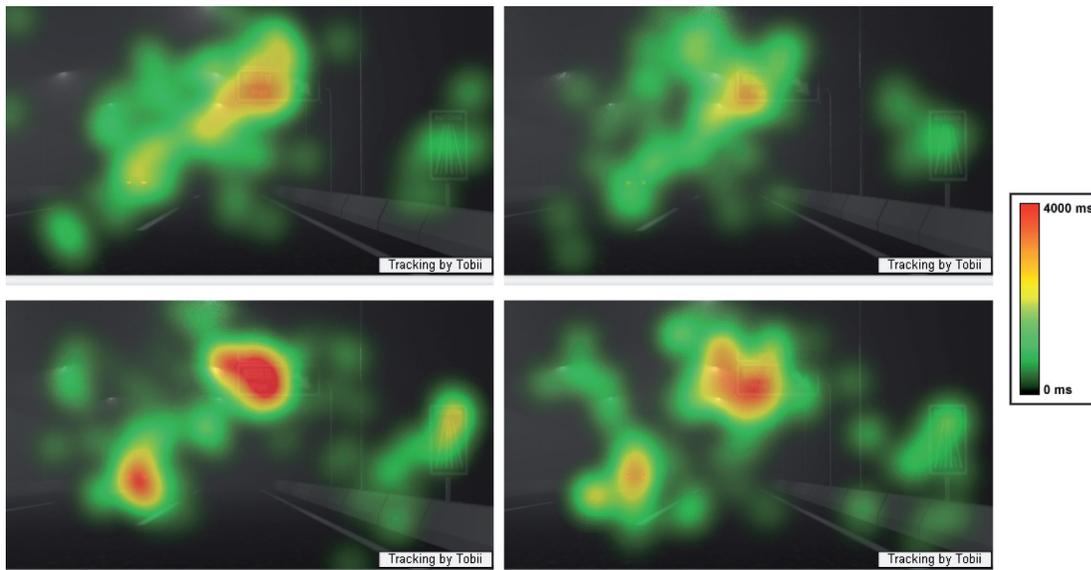


Fig. 15. Hotspot visualization of the total fixation durations (max value 4000 ms) for the 24 participants (no reference condition): 12 participants watched the images in the top pair consecutively, where (left) is a high-quality image and (right) rendered using the XS map. The other 12 participants watched the bottom pair where (left) is rendered using the XS map and (right) a high-quality image.

extending our previous XS map with the L map, we have shown experimentally that we can significantly reduce rendering time while maintaining a high perceptual result. The introduction of the L map allow us to overcome visible artifacts caused by light scattering in selective rendering of participating media.

6.5 Eye-Gaze Visualization

In addition to the subjective validation the Tobii x50 eye-tracking system [Tobii 2006] was used on the participants during the first experiment. Eye tracking allow us to study where the participant's gaze falls on the computer screen and, hence, what drew their attention. It can also give an indication as to how the observer perceived the environment. In general, the more attention an area has received, the more fixations are clustered here. The x50 system samples the position of the participant's eyes at a constant frame rate of 50 Hz and has an average accuracy of about $0.5\text{--}0.7^\circ$. A calibration was carried out for each participant, prior to each experimental trial, to ensure the collected data would be reliable. Each calibration took 30 s on average. One useful method in order to display graphically where a group of people was attending to is called a *hotspot visualization* [Tobii 2006]. The hotspot plot consists of the experiment stimuli as a background image with a hotspot mask superimposed on top. The hotspot mask consists of a black background, with high lighted areas, showing where the participants have been looking. A hotspot visualization of the total fixation duration of the 24 participants (in the no reference condition) was created for the road scene stimuli. The maximum value of the total fixation duration was set to 4000 ms, the same length of time as each image was displayed for. Figure 15 shows the hotspot plot for the Road scene. The longest durations seems to be in the areas which are also most salient as indicated in the saliency map in Figure 4 (top middle). The fixations are mostly gathered around both the traffic signs and the car lights. Longer fixations durations were also centered on the lights and the light scattering in the scene, which had a lot of artifacts around them because of the flaws in the XS map that later was corrected for by our proposed L map.

7. CONCLUSIONS AND FUTURE WORK

We have presented an original perceptual selective method for efficient physically based rendering of participating media. Within this technique, we have introduced the X map, which precomputes light attenuation in inhomogeneous participating media (with a straightforward simplification for the homogeneous case). Combined with a modified saliency map into the XS map, this map can be used to guide selective rendering of scenes including participating media, with high-quality antialiasing settings for salient features foreseen to be less attenuated by the medium. Using the XS map, we have been able to achieve a substantial speedup of 5–10 times depending on scene. For animated sequences such computational savings can, therefore, be highly significant.

Furthermore, we performed a perceptual validation of the selectively rendered stimuli generated using the XS map to assess the visual quality of the resulting images. The perceptual validation consisted of two parts: using two objective difference metrics (MSE and VDP) and a subjective psychophysical experiment. The average pixel errors were less than 1% for all scenes between a gold standard and our selectively rendered images. Although the average pixel error was low, the subjective experiments showed that participants still could detect a reduction in visual quality of selectively rendered images for certain scenes. Our experiment also extends previous work by showing that a low average pixel error does not necessarily mean that we can obtain a result that is perceptually indistinguishable from a full solution. In the absence of an ideal solution, it is still possible, however, to achieve a substantial reduction in rendering time while maintaining a high-quality perceptual result.

One limitation of our the X map and XS map was that they could not foresee strong scattering around light sources, which would become an important salient feature of the scene. The XS map was, therefore, extended with a novel light map (L map) to overcome problem areas that were sampled poorly, resulting in visible artifacts. The combination of the XS map with our new L map was termed the XSL map, which gives a modulated information about the salient features of the scene and the extinction and scattering produced by the present participating media. We also performed an additional perceptual validation using the new XSL map to assess the visual quality of the resulting images. Although we did not reduce the computational cost as much as by using the XS map, the experimental results showed that the XSL map can be used to maintain a high-fidelity perceptual result for scenes with visible light sources against an empty background.

For future research it would also be interesting to study perceptually based rendering of dynamic environments, including a participating media. Our current approach does not take into account color of participating media and thus potential salient areas could be missed. Other important issues that should be addressed are partially occluded light sources, currently overlooked by our Gaussian-based technique, and highlights, which can attract an observer's attention.

ACKNOWLEDGMENT

This research was sponsored by the Spanish Ministry of Education and Research through the project TIN2004-07672-C03-03 and the European Union within the CROSSMOD project (EU IST-014891-2). We would also like to thank Adrian Gomez and Kurt Debattista for their collaboration in the research and everyone that participated in the experiments. The model of Lucy was obtained from the Stanford University Computer Graphics Laboratory.

REFERENCES

- ANSON, O., SUNDSTEDT, V., GUTIERREZ, D., AND CHALMERS, A. 2006. Efficient selective rendering of participating media. In *APGV 06: Symposium on Applied Perception in Graphics and Visualization*. ACM Press, New York. 135–142.
- BOLIN, M. R. AND MEYER, G. W. 1995. A frequency based ray tracer. In *SIGGRAPH*. 409–418.
- ACM Transactions on Applied Perception, Vol. 4, No. 3, Article 15, Publication date: November 2007.

- BOLIN, M. R. AND MEYER, G. W. 1998. A perceptually based adaptive sampling algorithm. *Computer Graphics 32*, Annual Conference Series, 299–309.
- CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: exploiting visual tasks for selective rendering. In *EGRW '03: Proc. of the 14th Eurographics Workshop on Rendering*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland. 270–280.
- CHALMERS, A., DEBATTISTA, K., AND SANTOS, L. P. 2006. Selective rendering: Computing only what you see. In *Proceedings of GRAPHITE 2006, ACM SIGGRAPH*.
- DALY, S. 1993. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*. A.B. Watson, Ed. MIT Press, Cambridge, MA. 179–206.
- DEBATTISTA, K. AND CHALMERS, A. 2005. Component-based adaptive sampling. In *SIBGRAPI 2005*. IEEE Computer Society Press, Los Alamitos, CA. 375–382.
- DUMONT, R., PELLACINI, F., AND FERWERDA, J. A. 2003. Perceptually-driven decision theory for interactive realistic rendering. *ACM Trans. Graph.* 22, 2, 152–181.
- FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. 1997. A model of visual masking for computer graphics. *Computer Graphics 31*, Annual Conference Series, 143–152.
- GLASSNER, A. S. 1994. *Principles of Digital Image Synthesis*. Morgan Kaufmann, San Francisco, CA.
- GUTIERREZ, D., MUNOZ, A., ANSON, O., AND SERÓN, F. J. 2005. Non-linear volume photon mapping. In *Proc. of the Eurographics Symposium on Rendering Techniques, Konstanz, Germany, June 29–July 1, 2005*. 291–300.
- HABER, J., MYSZKOWSKI, K., YAMAUCHI, H., AND SEIDEL, H.-P. 2001. Perceptually guided corrective splatting. In *Computer Graphics Forum, Proc. of Eurographics 2001*, A. Chalmers and T.-M. Rhyne, Eds. Computer Graphics Forum, vol. 20. Eurographics, Blackwell, Manchester, UK. C142–C152.
- ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 11, 1254–1259.
- JENSEN, H. W. AND CHRISTENSEN, P. H. 1998. Efficient simulation of light transport in scenes with participating media using photon maps. In *SIGGRAPH 98 Conference Proceedings*, M. Cohen, Ed. Annual Conference Series. ACM SIGGRAPH, Addison Wesley, Reading, MA. 311–320.
- KAJIYA, J. T. 1986. The rendering equation. In *SIGGRAPH '86: Proc. of the 13th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 143–150.
- LONGHURST, P., DEBATTISTA, K., GILLIBRAND, R., AND CHALMERS, A. 2005. Analytic antialiasing for selective high fidelity rendering. In *SIBGRAPI 2005*. IEEE Computer Society Press, Los Alamitos. 359–366.
- LONGHURST, P., DEBATTISTA, K., AND CHALMERS, A. 2006. A gpu based saliency map for high-fidelity selective rendering. In *AFRIGRAPH 2006 4th Int. Conf. on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*. ACM SIGGRAPH, 21–29.
- MARTENS, W. AND MYSZKOWSKI, K. 1998. Psychophysical validation of the visible differences predictor for global illumination applications. In *Proceedings of IEEE Visualization '98*. 18–23.
- MITCHELL, D. P. 1987. Generating antialiased images at low sampling densities. In *SIGGRAPH '87: Proc. of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 65–72.
- MYSZKOWSKI, K. 1998s. The visible differences predictor: Applications to global illumination problems. In *Proc. of the Eurographics Workshop*. 223–236.
- MYSZKOWSKI, K. 2002. Perception-based global illumination, rendering, and animation techniques. In *SCCG '02: Proc. of the 18th Spring Conference on Computer Graphics*. ACM Press, New York. 13–24.
- NARASIMHAN, S. G. AND NAYAR, S. K. 2003. Shedding light on the weather. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. I. 665–672.
- NEUMANN, D. AND GEGENFURTNER, K. R. 2002. Perception based image retrieval.
- PRIKRYL, J. 2001. Radiosity methods driven by human perception. Ph.D. thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology.
- RAMASUBRAMANIAN, M., PATTANAIK, S. N., AND GREENBERG, D. P. 1999. A perceptually based physical error metric for realistic image synthesis. In *SIGGRAPH '99: Proc. of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley, New York. 73–82.
- RUSHMEIER, H. 1994. Rendering participating media: Problems and solutions from application areas. In *Proc. of the 5th Eurographics Workshop on Rendering*. 117–126.
- SIEGEL, R. AND HOWELL, J. R. 1992. *Thermal Radiation Heat Transfer*. Hemisphere Publ., New York.
- STOKES, W. A., FERWERDA, J. A., WALTER, B., AND GREENBERG, D. P. 2004. Perceptual illumination components: A new approach to efficient, high quality global illumination rendering. *ACM Trans. Graph.* 23, 3, 742–749.

- SUN, B., RAMAMOORTHY, R., NARASIMHAN, S. G., AND NAYAR, S. K. 2005. A practical analytic single scattering model for real time rendering. *ACM Trans. Graph.* 24, 3, 1040–1049.
- SUNDSTEDT, V., DEBATTISTA, K., AND CHALMERS, A. 2005a. Perceived aliasing thresholds in high-fidelity rendering. In *APGV '05: Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*. ACM Press, New York. 166–166.
- SUNDSTEDT, V., DEBATTISTA, K., LONGHURST, P., CHALMERS, A., AND TROSCIANKO, T. 2005b. Visual attention for efficient high-fidelity graphics. In *Spring Conference on Computer Graphics (SCCG 2005)*. 162–168.
- Tobii 2006. *User Manual: Tobii Eye Tracker; ClearView Analysis Software*. Tobii.
- WOOLLEY, C., LUEBKE, D., WATSON, B., AND DAYAL, A. 2003. Interruptible rendering. In *SI3D '03: Proceedings of the 2003 Symposium on Interactive 3D Graphics*. ACM Press, New York. 143–151.
- YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.* 20, 1, 39–65.

Received December 2006; revised April 2007; accepted May 2007