Efficient Selective Rendering of Participating Media

Oscar Anson*
University of Zaragoza
Spain

Veronica Sundstedt†
University of Bristol
United Kingdom

Diego Gutierrez‡
University of Zaragoza
Spain

Alan Chalmers§
University of Bristol
United Kingdom

Abstract

Realistic image synthesis is the process of computing photorealistic images which are perceptually and measurably indistinguishable from real-world images. In order to obtain high fidelity rendered images it is required that the physical processes of materials and the behavior of light are accurately modelled and simulated. Most computer graphics algorithms assume that light passes freely between surfaces within an environment. However, in many applications, ranging from evaluation of exit signs in smoke filled rooms to design of efficient headlamps for foggy driving, realistic modelling of light propagation and scattering is required. The computational requirements for calculating the interaction of light with such participating media are substantial. This process can take many minutes or even hours. Many times rendering efforts are spent on computing parts of the scene that will 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 inhomogenous media. We also validate the visual quality of the resulting images using two objective difference metrics and a subjective psychophysical experiment. Although the average pixel errors of these metric are all less than 1%, the experiment using human observers indicate that these degradation in quality is still noticeable in certain scenes, unlike previous work has suggested.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism-color, shading, shadowing, and texture—;

Keywords: participating media, perception, saliency map, extinction map, selective rendering

1 Introduction

Rendering physically-based imagery by using global illumination techniques have become very useful and necessary in certain areas of interest like safety, military or industry [Rushmeier 1994]. The applications in these areas usually deals with the analysis of visual perception under certain 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, road lighting, etc. Examples of these participating media include smoke, fog, dust,

---

*e-mail: oanson@unizar.es
†e-mail: veronica@cs.bris.ac.uk
‡e-mail: diegog@unizar.es
§e-mail: alan@cs.bris.ac.uk

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

flames, silty, and abyssal waters or atmospheric phenomena, Figure 1.

Simulating these participating media implies the correct computation of the absorption and scattering of light and therefore it has been always computationally expensive.

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), which stores in image-space the exponential decay of light in the medium. This combination, that we have called the XS-map, is then used to guide a selective rendering of the scene, with more accurate estimates in the most perceptually important areas of the scene. The novelties of our work can then be summarized as the introduction of the X-map concept and its combination with a saliency map to guide a perception-based renderer for inhomogeneous participating media. In addition, two objective metrics are used in order to validate our approach and the perceived quality of selective rendering results is assessed by a two forced-choice preference experiment.

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 XS-map in Section 4 and present the results in Section 5. Finally, we validate the XS-map using both objective metrics and subjective experiments in Section 6. 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

Figure 2: The three maps computed for the Road scene in Figure 9: the X-map (left), the saliency map (center) and the XS-map (right).

based on visual attention had been developed before, such as adaptive sampling [Mitchell 1987] for ray tracing which applied perceptual considerations, they have become more popular recently. An excellent overview of early perceptually-driven radiosity methods is given in [Prikryl 2001]. A frequency based ray tracer using a more complete model of the human visual system which incorporated the visual systems spatial processing behaviour and sensitivity change as a function of luminance was developed by [Bolin and Meyer 1995]. An even more sophisticated model incorporating visual 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 Dalys 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 keyframe based animation sequence. Myskowskis 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. 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; Sundstedt et al. 2005] both task maps and saliency map were used to vary a number of rays shot per pixel in a global illumination environment.

3 Modeling Participating Media

Traditional Radiance Equation introduced in computer graphics by [Kajiya 1986] and solved by global illuminations algorithms can only be used if vacuum is the only medium in the scene, where the interaction of light is completely non-existent. But when rendering scenes that contain participating media, the interaction between the light and particles floating in the medium has to be taken into account. A physically accurate lighting simulation of such media implies solving the Radiative Transfer Equation (RTE) [Glassner 1994], which is an integro-differential equation and noticeably more complex. Different resolution strategies are surveyed in [Perez et al. 1997].
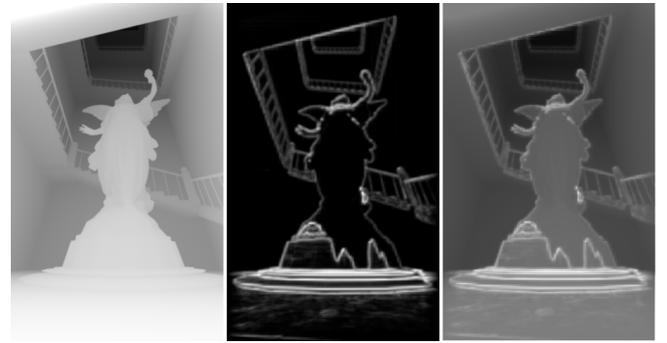
Figure 3: The three maps computed for the Lucy scene in Figure 9: the X-map (left), the saliency map (center) and the XS-map (right).

The RTE models the three basic effects where the light passes through a participating medium, each of them governed by a particular wavelength-dependent coefficient. These three types of interaction are emission, absorption and scattering (in-scattering and out-scattering). The RTE describes the variation of radiance L_λ in a point x in the direction ω and 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 coefficient, and $L_{e,\lambda}$ and $L_{i,\lambda}$ are the emitted and in-scattered radiance respectively. As it can be seen, both absorption and scattering coefficients vary throughout the medium. In that general case, the medium is named inhomogeneous, if both coefficients are constant the medium will be homogeneous.

Defined the extinction coefficient as $\kappa_\lambda(x) = \alpha_\lambda(x) + \sigma_\lambda(x)$ and knowing that the in-scattered radiance $L_{i,\lambda}$ depends of the radiance incoming from all possible directions $\vec{\omega}$ into the point x over the sphere Ω we can rewrite the Equation 1 in the form

$$\frac{\partial L_\lambda(x,\vec{\omega})}{\partial x} = \alpha_\lambda(x)L_{e,\lambda}(x,\vec{\omega})$$

$$\begin{aligned}
+ \quad & \sigma_\lambda(x) \int_\Omega p_\lambda(x, \vec{\omega}', \vec{\omega}) L_\lambda(x, \vec{\omega}') d\vec{\omega}' \\
- \quad & \kappa_\lambda(x) L_\lambda(x, \vec{\omega}) \qquad\qquad\qquad\qquad (2)
\end{aligned}$$

where $p_\lambda(x, \vec{\omega}', \vec{\omega})$ 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 Equation 2 by integrating with respect its boundary conditions we have the integral form of the RTE [Siegel and Howell 1992] given by

$$\begin{aligned}
L_\lambda(x, \vec{\omega}) \;=\; & e^{-\tau_\lambda(x_0, x)} L_\lambda(x_0, \vec{\omega}) \\
+ \; & \int_{x_0}^{x} e^{-\tau_\lambda(x', x)} \alpha_\lambda(x') L_{e,\lambda}(x', \vec{\omega}) dx' \\
+ \; & \int_{x_0}^{x} e^{-\tau_\lambda(x', x)} \sigma_\lambda(x') \int_\Omega p_\lambda(x', \vec{\omega}', \vec{\omega}) L_\lambda(x', \vec{\omega}') d\vec{\omega}' dx'
\end{aligned}$$
$$(3)$$

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

$$\tau_\lambda(x', x) = \int_{x'}^{x} \kappa_\lambda(s) ds \qquad\qquad (4)$$

As it can be seen in Equation 3, the computation of the radiance at a point in the scene directly depends on the radiance of all other points in the medium around it. That fact makes the computation for scenes with participating media more complex and extremely computationally expensive than the traditional radiance equation; even more if Equation 3 is extended with special wavelenght-dependant effects such as inelastic scattering [Gutierrez et al. 2005]. In the simplest case, where the medium is homogeneous, the Equation 3 could be simplified.

4 Selective rendering of Media

In order to selectively render nonhomogenous participating media, we have extended our in-house rendering system named *Lucifer* [Gutierrez et al. 2005] with a perceptually-based director. The overall goal is to compute scenes with physically accurate participating media, in a reasonable time, while maintaining a high perceptual result. This process is completed in two stages. The first stage consists of the generation of an extinction map (X-map) and a saliency map and their combination into the directing map (XS-map). In the second stage, the XS-map, that is a gray scale image, is then used to drive the rendering process itself, rendering in higher quality the brighter areas of the XS-map.

4.1 The X-Map

In the general case of inhomogeneous participating media, the intensity of light as it transverses the medium is reduced by the transmittance $e^{-\tau_\lambda(x', x)}$. This attenuation in the optical path length τ_λ between x' and x described in Equation 4 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 human eye. The first idea of this work is to pre-compute that fraction beforehand, storing in image-space attenuation values in an extinction map which we have named the X-map.

To obtain the X-map, rays are casted from the eye into the scene, ray marching through the medium until the first intersection, and save the result of the exponential decay in the X-map, representing the percentage of light (both direct and diffuse) which will reach the eye for each pixel (or sub-pixel if the resolution of the map is increased). For homogeneous media, the attenuation is only $e^{-\kappa_\lambda s}$, 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 instantaneous recomputations of the X-map if the description of the medium in the scene changes, since intersections do not need to be recalculated at each step.
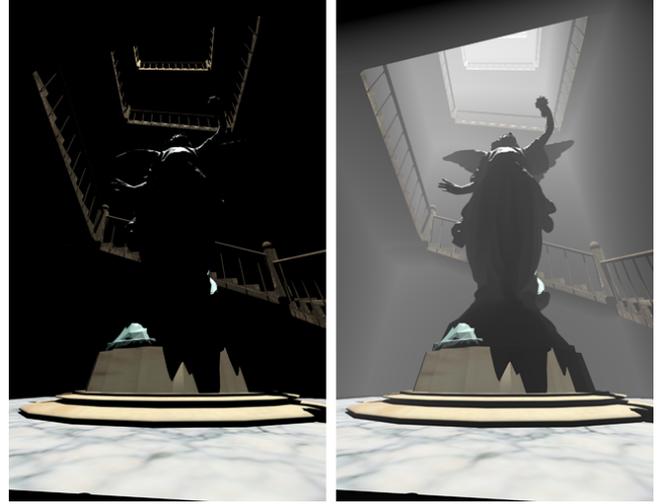
Figure 4: Fast snapshots for the Lucy scene. Left: initial fast raytraced snapshot (direct lighting and no medium). Right: A false estimate of the medium computed from the combination of the raytraced snapshot and the inverted X-map.

After generating the X-map, a fast raytraced snapshot of the scene (without the medium) is computed as well [Longhurst et al. 2005; Sundstedt et al. 2005]. This step is completed in the order of few seconds because only direct lighting is computed (see Figure 4 left), and does not add any significant overhead to the overall process time. But the main aim is to obtain a rapid estimate of the scene with the medium, which will be used as an input to a saliency generator in order to detect the features of the image that attract the attention of the observer when the medium is present in the scene. But as it was seen in Section 3, rendering the medium can be computationally expensive. Instead that, we used the computed X-map, inverting its values, in combination with the early fast snapshot obtaining an estimate with a false medium. The result is shown in Figure 4 right.

In that sense, the X-map is used in two ways. Firstly for simulating the extinction in the scene due to the medium, and secondly, to combine it with the output of the saliency generator.

4.2 The Saliency Map

The saliency generator, that is a bottom-up, feature-based model, extracts features from an image that humans would automatically direct attention to. That input image, in our case, is a fast estimate of the scene generated from the combination of the fast snapshot and an inverted X-map.

The saliency map is then computed from this estimate (in 2-3 seconds per image) via a combination of three conspicuity maps of

intensity, colour 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. A hardware implementation can generate a saliency map in the order of tens of milliseconds [Longhurst et al. 2006].

4.3 The XS-Map

To complete our setup for selective rendering of participating media, we further propose a novel combination of this X-map with the saliency map, based on the work of [Sundstedt et al. 2005]. This combination is given by $XS(w_x, w_s, op)$, where XS represents the combined saliency and X- maps, and w_x and w_s are coefficients applied to the values in the X-map and the saliency map respectively, which allow 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. Our results in Section 5 were computed using addition to combine the information from both maps such that all weighted features are preserved. Such equal weighting would 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 represents 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.

Figures 2 and 3 show the various maps for one of our test scenes, as shown in Figure 9. Figures 2 and 3 (left) show the X-map. Figures 2 and 3 (center) demonstrate the saliency map. Figures 2 and 3 (right) show the combination of the X-map and the saliency map with equal weighting into the XS-map.

5 Results

We have implemented our presented method in *Lucifer* renderer system that now uses the XS-map to direct some rendering parameters. 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 XS-map. Higher values in this director map will result in lower aliasing effects and therefore increased quality for the given pixels. 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 four different scenes to construct the tests for our perceptual validation. We name them as Balls A, Balls B, Lucy and Road. For each scene, as shown in Figure 9, we rendered four versions: the gold standard, using only the saliency map, using only the X-map and, and finally, using the XS-map. In the gold standard image, sixteen rays per pixel where shot for each pixel in the whole image, giving a high quality solution reference. For the three remaining versions, a sixteen rays per pixel ceiling was established.

The whole set of sixteen 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 5 shows the timing comparison between the reference gold standard and the solution generated using the
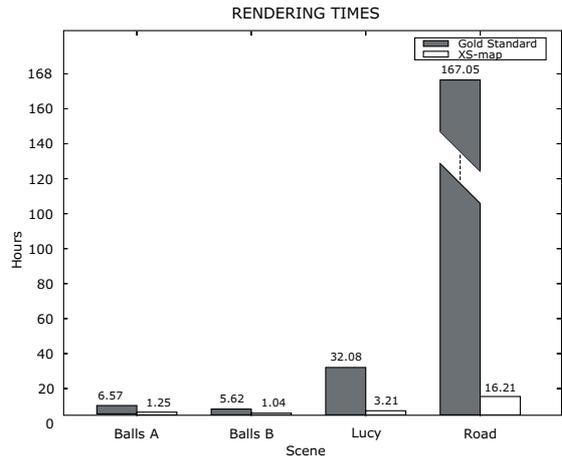
Figure 5: Timing comparison for the four scenes between the reference gold standard solution and the image generated using the XS-map.

XS-maps. Rendering gold standard images takes between 5.2 and 10.3 times longer than using our XS-map.

Although by rendering scenes with the saliency map is approximately 1.15 times faster than using the XS-map, the visual quality that we get with the latter allows us to afford that extra rendering time. Figure 6 shows a comparison between both images.

6 Perceptual Validation

In order to measure the visual quality of the selectively rendered solutions, two techniques were used. Firstly, two objective difference metrics were used to find the average pixel error between the gold standard for each scene and their corresponding XS-map renderings. Secondly, a subjective psychophysical validation was performed, using a two-alternative forced-choice (2AFC) preference experiment.

6.1 Objective Difference Metrics

The first difference metric used was the mean square error (MSE). Using MSE a difference map could be calculated which then was averaged for all pixels. The second metric used was Daly's perceptually-based visual differences predictor (VDP) [Daly 1993]. The VDP generates a pixel-by-pixel map of the probability that, at any given pixel in the image, a human viewer will perceive a difference between two images. This 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. For example, the same average could refer to an imperceivable error throughout the image or a highly visible error in a certain location.

The MSE and VDP average percentage pixel error results can be seen in Table 1. Both average pixel error values were less than 1% for each of the four scenes. Another way to display the results are the percentage of pixel errors for each scene, Table 2. All VDP pixel error values were also 0.57% or less. The MSE pixel error 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.

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

Table 2: MSE and VDP pixel error between the four scenes gold standard and the XS-map renderings.

played at the same time, Figure 7.

6.2.1 Participants

48 participants took part in the experiment (42 men and 6 women; age range: 19-38). 24 participants took part in each condition. Subjects had a variety of experience with computer graphics, and all self-reported normal or corrected-to-normal vision.

6.2.2 Stimuli

The visual trial in the experiment was based upon four scenes, namely: Balls A, Balls B, Lucy, and Road, 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. 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 centre of the screen, at a viewing distance of 60 cm. All stimuli were displayed on a screen which had a 50% grey background.

6.2.3 No Preference 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 four seconds each with a 50% grey image in between lasting for two seconds. 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. Half the high quality images were on the left and half on the right, in random order. The first condition was chosen to study if 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.

6.2.4 Preference Procedure

In the second condition participants had to discriminate, using a 2AFC method, 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 XS-map rendered image and the gold standard were displayed below, side by side, Figure 7. The three images were displayed during a time of eight seconds. Half of the gold standard images were on the left and half on the right, in random order. After each trial, participants

Figure 6: Comparison between the image generated only using the saliency map (left) and the one rendered with our XS-map (right). On bottom, a zoomed region shows the aliasing differences between them.

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

Table 1: Average MSE and VDP pixel error between the four scenes gold standard and the XS-map renderings.

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 judgements in evaluating progressive radiosity solution convergence [Martens and Myszkowski]. Even though the average pixel errors for all scenes were less than 1%, a psychophysical experiment was run with 48 human observers. In this experiment participants performed a 2AFC, that assessed the perceived quality of selective rendering using the XS-map. Within this experiment two different display conditions for the experimental images were used, Section 6.2.3 and 6.2.4:

No preference: discriminating which of two consecutively displayed images (gold standard and XS-map rendered in altered order) contained the worse rendering quality using a 2AFC task.

Preference: discriminating which of two images (gold standard and XS-map rendered in altered order) were most similar to a gold standard. All three images in this condition were dis-

Figure 7: 2AFC display used in the preference condition. At the top the reference image (gold standard) is shown. Below are the two experiment images (gold standard and XS-map rendered image were randomized with half on each side.

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 most fair 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 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 8 shows the overall results of the 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 towards one case, and indicates that no differences between the gold standard and the XS-map rendered images were perceived. For the first condition, 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.

6.4 Statistical Analysis and Discussion

The results were analysed 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 appropriate analysis is a one-sample chi-square test due to its nonparametric nature. The chi-square values were computed and then tested for significance, Table 3.
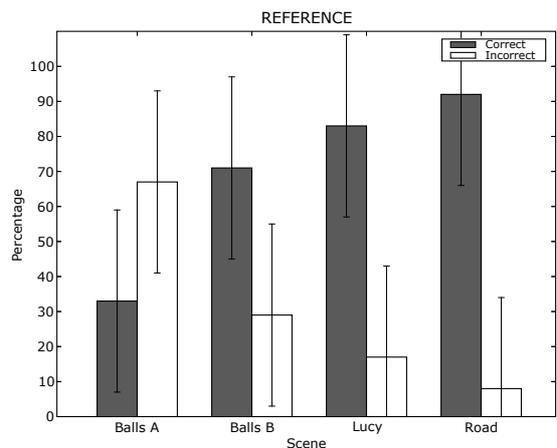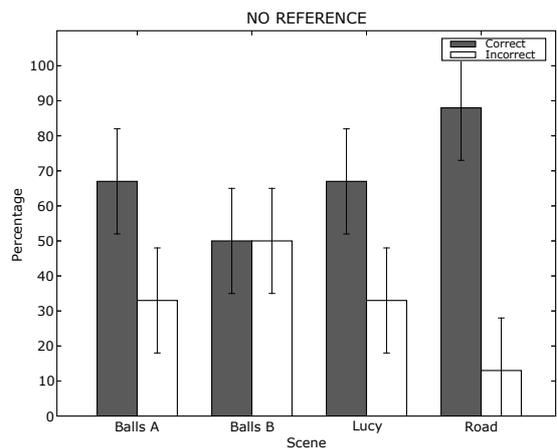
Figure 8: Experiment results for the two conditions: no reference vs. reference with error bars representing standard deviation.

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 ($p > 0.5$), Table 3. This indicates that the participants not were able to correctly classify the image with worse rendering quality. For the Road scene the result were highly significant ($p < 0.5$). 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 ($p > 0.5$). This indicates that the participants not were 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 ($p < 0.5$). 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

	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	$\chi^2(1, N = 24) = 13.5, p = 0.00$	$\chi^2(1, N = 24) = 16.67, p = 0.00$

Table 3: Output for a Chi-square Analysis.

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 show the importance of using human observers in the validation process of realistic image synthesis algorithms.

7 Conclusions and Future Work

We have presented an original selective rendering system for efficient physically-based simulation of participating media. Within this system, we have introduced a novel concept, the X-map, which precomputes light attenuation in inhomogeneous participating media (with a straightforward simplification for the homogeneous case). Combined with a 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 the computational savings can therefore be highly significant.

Furthermore, we performed a perceptual validation of the selectively rendered stimuli to assess the visual quality of the resulting images. The perceptual validation consisted of two parts: using two objective difference metrics and a subjective psychophysical experiment. On average the resulting pixel error 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 extend previous work by showing that a low average pixel error not necessarily means that we can obtain a result indistinguishable from a full solution. Although, in the absence of an ideal solution it was possible to achieve a substantial reduction in rendering while keeping a high perceptual result.

Future work will look at the extension of the X-map by supporting other effects due to the participating media such scattering of light. This will be important in order to predict bright areas around light sources that are not supported in our work, as it can be seen in the Road scene, Figure 9.

8 Acknowledgements

References

BOLIN, M. R., AND MEYER, G. W. 1995. A frequency based ray tracer. In *SIGGRAPH*, 409–418.

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.

DALY, S. 1993. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity. In *Digital Images and Human Vision*, A.B. Watson, MIT Press, Cambridge, MA, 179–206.

DEBATTISTA, K., AND CHALMERS, A. 2005. Component-based adaptive sampling. In *SIBGRAPI 2005*, IEEE Computer Society Press, 375–382.

DEUSSEN, O., KELLER, A., BALA, K., DUTRÉ, P., FELLNER, D. W., AND SPENCER, S. N., Eds. Proceedings of the Eurographics Symposium on Rendering Techniques, Konstanz, Germany, June 29 - July 1, 2005.

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 Publishers Inc., San Francisco, CA, USA.

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*, Blackwell, Manchester, UK, A. Chalmers and T.-M. Rhyne, Eds., vol. 20 of *Computer Graphics Forum*, Eurographics, 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.

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, NY, USA, 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, 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. Psychophysical validation of the visible differences predictor for global illumination applications. In *IEEE Visualization '98 (Late Breaking Hot Topics)*, 49–52.

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*, 65–72.

MYSZKOWSKI, K. 1998. 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, NY, USA, 13–24.

PEREZ, F., PUEYO, X., AND SILLION, F. X. 1997. Global illumination techniques for the simulation of participating media. In *Proc. of the Eurographics Workshop on Rendering Techniques '97*, Springer-Verlag, London, UK, 309–320.

PRIKRYL, J. 2001. *Radiosity methods driven by human perception*. PhD 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*, 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 Publishing Corporation, 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.

SUNDSTEDT, V., DEBATTISTA, K., LONGHURST, P., CHALMERS, A., AND TROSCIANKO, T. 2005. Visual attention for efficient high-fidelity graphics. In *Spring Conference on Computer Graphics (SCCG 2005)*, 162–168.

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.
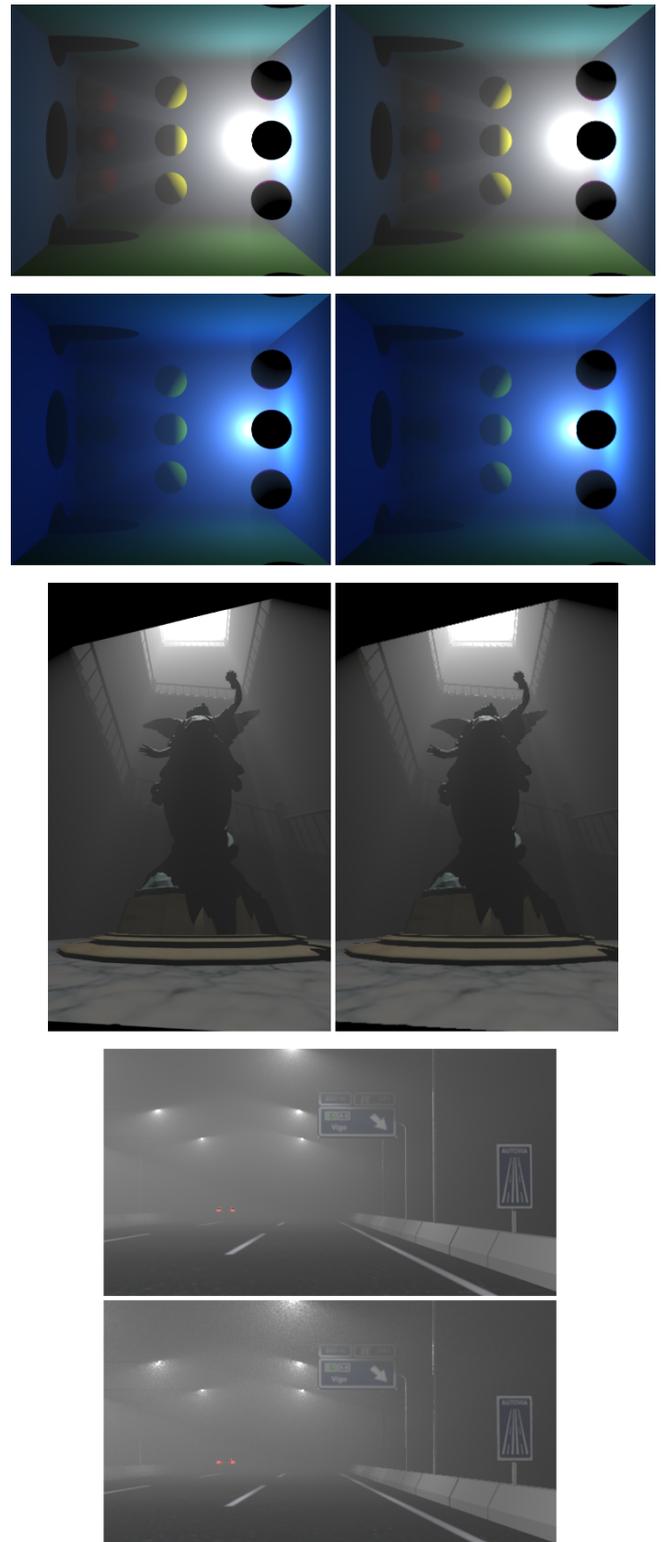
Figure 9: Stimuli for the visual trial. From top to down: Balls A, Balls B, Lucy and Road. On the left side, the gold standard images are shown. On the right side, the XS-map images are shown. In the case of the Road scene, the gold standard is on top.