

Perception-Guided Global Illumination Solution for Animation Rendering

Karol Myszkowski

Takehiro Tawara

Hiroyuki Akamine

Hans-Peter Seidel

Max-Planck-Institut für Informatik*

Abstract

We present a method for efficient global illumination computation in dynamic environments by taking advantage of temporal coherence of lighting distribution. The method is embedded in the framework of stochastic photon tracing and density estimation techniques. A locally operating energy-based error metric is used to prevent photon processing in the temporal domain for the scene regions in which lighting distribution changes rapidly. A perception-based error metric suitable for animation is used to keep noise inherent in stochastic methods below the sensitivity level of the human observer. As a result a perceptually-consistent quality across all animation frames is obtained. Furthermore, the computation cost is reduced compared to the traditional approaches operating solely in the spatial domain.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Time-varying imagery

Keywords: Animation, Human Factors, Illumination, Monte Carlo Techniques, Temporal Aliasing

1 Introduction

Global illumination is an important visual cue, which greatly improves the appearance of rendered images. In this work we focus on high quality computer animations that are becoming more and more widespread in many applications such as entertainment, advertisement, education, engineering, architecture, urban planning, and many others. Obviously, adding global illumination effects to such animations enhances their realism and makes them more believable in reconstructing the appearance of the real world.

Although the off-line computation of top quality computer animations makes it possible to include costly lighting simulation techniques, it almost never happens in industrial practice [1]. The main problem with existing global illumination solutions is poor scaling of the computation load with increasing scene complexity, which is often caused by wasting computational efforts on unimportant scene details that cannot be perceived in the final animation by the human observer [23, 11]. Also, existing animation rendering techniques process every single frame independently, and therefore cannot account properly for the eye sensitivity variations resulting

from temporal considerations [18]. While some efficient techniques of perception-based guidance of the global illumination computation for static images have been proposed recently [7, 3, 22], relatively little attention has been paid to the temporal aspect of such a guidance for dynamically changing environments. In particular, indirect lighting often changes slowly from frame to frame and many of those changes might be too subtle to be perceived (changes in direct lighting are usually quite well noticeable due to high-contrast shadows and highlights accompanying the moving objects). Since it would be an obvious waste to repeat the computation of indirect lighting for every frame from scratch, the performance of animation rendering could be significantly improved by exploiting the temporal coherence of indirect lighting. Ideally, lighting information acquired for the preceding and following frames should be re-used as much as possible for reconstructing indirect lighting in a given frame; however, the quality of the resulting animation as perceived by the human observer should not be compromised.

Some attempts were made to exploit the coherence in lighting distribution by repeating the indirect lighting computation for a fixed number of frames (so-called *keyframes*) and re-using the results obtained for the remaining frames (so-called *inbetween frames*). Usually the number of inbetween frames between a pair of keyframes is the same for the whole animation, and there is no control of the validity of applying the keyframe lighting to the inbetween frames [31]. Such an approach might result in visually noticeable errors in the lighting distribution, which is affected by changes in the environment that occur in the course of the animation. Obviously, the errors in lighting that are explicitly caused by the scripted animation of light sources can be compensated by increasing the number of keyframes for the affected animation segments. However, the question arises how many additional keyframes must be placed, so that approximations in the lighting reconstructed for inbetween frames remain unnoticeable. It is even more difficult to predict how the moving objects will affect the lighting distribution based merely on the animation script. Clearly, even for the simple approach with re-using lighting for inbetween frames some error metrics are needed to guide the keyframe placements. Ideally, some perception-based animation quality metrics are required that can directly judge whether the errors introduced by exploiting the temporal coherence are below the sensitivity level of the human observer. Also, by performing some limited computation for all frames (not just keyframes) abrupt changes in lighting can be identified more easily.

In this paper we introduce a novel framework for efficient global illumination computation in dynamic environments. We propose a combination of energy- and perception-based error metrics which efficiently guide lighting computation. Using these metrics the spatiotemporal coherence in the lighting distribution can be better exploited and the accuracy of the lighting computation can be relaxed without degradation of the animation quality as perceived by the human observer. As a result a perceptually-homogeneous quality of indirect lighting reconstruction across the spatial and temporal domains is obtained.

The paper is organized as follows. Section 2 reviews the previous approaches to global illumination in dynamic environments and perception-based animation. In Section 3 we introduce our new framework. The outline of our global illumination solution

*MPI für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, {karol,tawara,akamine,hpseidel}@mpi-sb.mpg.de

and a description of the extensions needed for handling dynamic environments are discussed in Section 4. In Section 5 we present the spatiotemporal processing guided by a hybrid of energy- and perception-based error metrics. We discuss the obtained results using our techniques in Section 6. Finally, we conclude the paper and propose some future directions for this research.

2 Previous Work

In our discussion of previous work we will concentrate on the two essential issues most relevant to our research. At first, we briefly overview existing solutions for handling global illumination reconstruction for dynamic environments. In particular, we focus on the problem of indirect lighting computation. In the second part we discuss the perception-based guidance of such a computation, which mostly remains an unexplored problem.

2.1 Global Illumination Solutions in Dynamic Environments

The problem of global illumination for dynamically changing environments has attracted significant attention in the research community, and a number of solutions that focus mostly on interactive applications have been proposed. Two basic strategies have been explored: brute-force computation that is performed on the fly for every frame usually with the support of graphics hardware and knowledge-based computation focused on the scene regions affected by dynamic changes in the environment.

The rapidly improving speed and functionality of graphics hardware makes it possible to render images displaying advanced lighting effects (refer to [13] for a complete survey of such techniques). However, the explicit global illumination computation is usually not performed, and although the resulting images look believable, they poorly predict the appearance of the real world. A notable exception is the instant radiosity technique [16], which can handle dynamic environments by replacing a number of aged light paths by new paths in every frame and combining the resulting images in the accumulation buffer. Since the technique is view-dependent, the objects newly appearing in the field of view might be under-sampled before a sufficient number of light paths involving those objects have been accumulated. It might be difficult to handle objects outside the field of view but visible in mirrors.

The extension of radiosity techniques to handle dynamic environments proved to be quite successful. Early solutions [4, 8, 17] were embedded into the progressive radiosity framework and relied on shooting the corrective energy (possibly negative) to the scene regions affected by the environment changes. Much better performance was obtained for more recently introduced techniques that are based on hierarchical radiosity [6, 24, 5, 21]. A line-space hierarchy proposed by Drettakis and Sillion [6] enables fast identification of links affected by a scene modification, and leads to image updates at interactive rates for moderately complex environments. However, the memory requirements inherent in this technique are extremely high because, apart from storing the active links used for energy gathering, also passive (refined) links and shafts for the entire scene are stored. The problem of storing shafts was recently reduced by Schoeffel and Pomi [24], who store shafts only locally for regions affected by geometry changes. Damez and Sillion [5] explicitly incorporated time in the hierarchical radiosity framework and showed substantial improvements in the rendering performance of animated sequences. However, this was achieved at the expense of a significant increase of memory requirements which become impractical for complex scenes. Pueyo et al. [21] proposed a radiosity algorithm which is focused on exploiting the temporal coherence of subsequent animation frames for static camera parameters. All

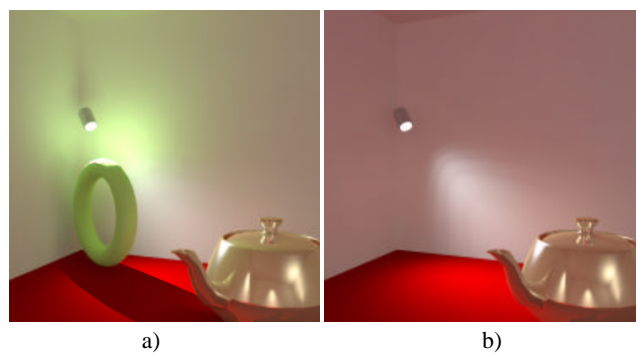


Figure 1: Indirect lighting changes can be significant in dynamic environments. Note the color bleeding effect in the room corner caused by the strongly illuminated torus in a). The effect disappears as the torus moves away in b).

discussed radiosity techniques only work well when the lighting changes are well localized in the scene. Otherwise, recomputing the lighting from scratch is a better choice. Also, the high quality of the lighting reconstruction is not guaranteed in those mesh-based object space approaches. In particular, changes to direct lighting effects such as shadows and highlights resulting from object motion are usually highly visible because they are important visual cues (in general specular effects are difficult to model for radiosity techniques). To avoid similar problems a majority of high quality rendering solutions computes direct lighting for every frame using view-dependent techniques such as sophisticated scanline methods [1], or ray tracing [30, 14]. We follow this strategy in our solution, thus we are more interested in techniques aiming at the efficient computation of indirect lighting in dynamic environments.

Changes in indirect lighting are usually slow and smooth. However, for dynamic environments such changes can occasionally be significant (refer to Figure 1). Many practical solutions ignore this possibility and compute the indirect lighting only for a limited number of frames, while reusing the obtained results for the remaining frames. This may cause some popping effects when the indirect lighting that was computed a number of frames backward is updated. To reduce this effect the indirect lighting should be sampled more densely, e.g., every 10 frames as in [31]. It is also possible that significant light changes are missed, or that periodic light changes are not properly captured. Clearly, some robust guidance for the selection of frames that require recomputation of indirect lighting is necessary.

A significant step in this direction was done by Nimeroff et al. [19] who proposed a powerful range-image based framework for handling indirect lighting in dynamic environments. The indirect lighting is sparsely sampled in time and then interpolated to reconstruct full global illumination for selected base images. The time steps for recomputing the indirect lighting are found by recursive subdivision. At each time step the lighting is calculated for a number of vertices using wavelet radiosity, then the differences between the corresponding vertices are computed. If differences larger than an assumed threshold are found for a certain percentage of vertices the time sequence is subdivided. The drawback of this approach is that direct lighting is not considered, which could effectively wash out even significant differences in indirect lighting [9]. Also, the tone reproduction [26] is not applied to the resulting lighting which is difficult in the view-independent framework as proposed by Nimeroff et al. [19] because the eye adaptation conditions cannot be established. Both effects can significantly affect the visibility of changes in indirect lighting and we take them into account in our technique.

The interpolation of indirect lighting between two time steps is

an important feature of Nimeroff’s framework. The continuity of changes in the lighting distribution between time steps is modelled and popping effects resulting from switching between two distinct lighting distributions as in [31] can be avoided. However, in all discussed approaches the accuracy of indirect lighting reconstruction fluctuates between frames, achieving the highest level for the keyframes, and then gradually decreasing for the remaining frames usually as a function of their distance to the keyframes along an animation path. In this research we investigate a novel approach in which sparse sampling of indirect lighting is performed for every frame, and the final lighting reconstruction is based on some processing of lighting distributions for a number of subsequent frames placed along the animation path. Thus, at each moment of time a similar level of accuracy of indirect lighting can be obtained. Such a framework is less prone to perceivable errors, and the probability of overlooking some important lighting events between keyframes is substantially reduced.

Since lighting samples collected in the time domain may become invalid for dynamic environments, a perception-based guidance of the lighting reconstruction is needed in order to reduce the probability of the perceivable errors resulting from such a reconstruction. In the following section we discuss previous work on perception-based solutions applied in the context of high quality animation.

2.2 Perception-Guided Animation Rendering

The main goal of perception-guided animation rendering techniques is to save computation without compromising the resulting animation quality as perceived by the human observer. Recently some successful examples of perception-based rendering of static images have been presented [7, 3, 22], however, expanding those techniques to handle the temporal domain remains mostly an open problem.

Myszkowski et al. [18] used a hybrid ray tracing and Image-Based Rendering (IBR) approach to improve the rendering performance by deriving as many pixels as possible, using inexpensive IBR techniques without affecting the walkthrough animation quality (the technique is not suitable for dynamic environments). A perception-based spatiotemporal Animation Quality Metric (AQM) was used to automatically guide such a hybrid rendering. The central part of the AQM is a model for the spatiovelocity Contrast Sensitivity Function (CSF), which specifies the detection threshold for a stimulus as a function of its spatial and temporal frequencies. Also, visual masking is modelled, which affects the detection threshold of a stimulus as a function of the interfering background stimulus which is closely coupled in space. The AQM models temporal and spatial mechanisms (channels) which are used to represent the visual information at various scales and orientations in a similar way as the primary visual cortex does. In this research we use the AQM as well but in the completely different context of guiding the global illumination computation rather than the IBR processing.

Since gains by adding further extensions to current early vision models such as the AQM are rather small [20], some attempts of using higher level perceptual and cognitive elements have been introduced in the context of animation. Yee [31] proposes an interesting application of a visual attention model to improve the efficiency of indirect lighting computations in the RADIANCE system [30] for dynamic environments. Yee demonstrated that greater errors can be tolerated for less salient image regions in which the density of indirect lighting samples can be substantially reduced. However, variability in the selection of the region of interests (ROI) for different observers, or even for the same observer from session to session, can lead to some degradation of the animation quality in regions that were not considered as important attractors of the visual attention. Yee reports that such degradations of quality could

be perceived when the same animation sequence was viewed more than once by the same observer. In our research, we are aiming at applications that require high quality animations which will possibly be viewed many times by a large number of observers. For this reason, we decided not to include visual attention models into our framework. Yee also ignored visual masking which plays an important role in hiding imperfections of reconstructed lighting [7].

3 Overview

The goal of this work is to improve the performance of global illumination computations for animated sequences of high quality by exploiting the temporal coherence in indirect lighting distribution in a better way. The mesh-based view-independent Density Estimation Particle Tracing (DEPT) algorithm [27], which we extend in this work to handle animated objects, is used as a global illumination framework, but the proposed solutions could be easily applied to other stochastic algorithms such as the photon map [14]. Initially, the lighting function is sparsely sampled in space for all frames (not just for fixed keyframes as in [19, 31]) within a given animation segment. Then, based on the obtained results, the decision is made whether the segment can be expanded/contracted in the temporal domain. Since the validity of samples may depend on the particular region in the scene for which indirect lighting conditions change more rapidly, different segment lengths are chosen locally for each mesh element (used to store particle hits), based on the variations of the lighting function. Energy-based statistical measures of such local variations are used to calculate the number of preceding and following frames for which samples can be safely used for a given region. More samples are generated if the quality of the frames obtained for a given segment length is not sufficient. The perception-based Animation Quality Metric (AQM) [18] is used to choose an average number of photons per frame for each segment to prevent perceivable degradation of animation quality. Spatial filtering is performed for those scene regions in which a sufficient number of samples cannot be collected in the temporal domain. For the final rendering the indirect lighting is reconstructed using the techniques just outlined, while specular effects and direct lighting are computed for every frame separately by ray tracing.

4 Indirect Lighting Solution

As a framework for global illumination computation, we chose the Density Estimation Photon Tracing (DEPT) algorithm [27]. The DEPT is similar to other stochastic solutions in which photons are traced from light sources towards surfaces in the scene, and the lighting energy carried by every photon is deposited at the hit point locations on those surfaces [12, 25, 29]. A simple photon bucketing on a dense triangular mesh is performed, and every photon is discarded immediately after its energy is distributed to the mesh vertices. Efficient object space filtering substantially reduces visible noise, while the excessive smoothing of the lighting function can be avoided by adaptively controlling the local filter support which is based on stochastically-derived estimates of the local illumination error [27, 29].

An important feature of the DEPT technique is that the bucket-based lighting reconstruction and filtering are very efficient, and the quality of the reconstructed lighting is quite good. Thus, the resulting illumination maps can be displayed immediately, and many variants of lighting reconstruction in the context of temporal photon processing for animations can be inexpensively analyzed for choosing the best solution. This is in contrast with other photon tracing techniques which involve costly density estimation techniques such as the kernel methods [25, 29], or the final gather step [14] in order to obtain images of good quality.

Another advantage of the DEPT computation is that a reasonable mesh-based approximation of the direct lighting is available (although in the final frames the direct lighting is recomputed on a per pixel basis), which is required to model the local eye adaptation conditions. The eye adaptation characteristics are needed for proper tone reproduction [26] during the illumination map display, and for the prediction of the eye sensitivity to the errors in indirect lighting. The reduction of these errors is the main objective of the perception-based guidance of temporal photon processing.

The extension of the DEPT algorithm to handle animated sequences of dynamically changing environments requires the storage of photons that are reused in neighbouring frames. The photons are stored on a per frame basis. For each photon, information on its spectral energy distribution is stored to account for reflected light and allow effects such as color bleeding (refer to Figure 1). Also, the hit point coordinates in the form of two barycentric coordinates are stored to facilitate distributing the photon energy to the mesh vertices, and to keep the photon position within the mesh elements in relative rather than in absolute terms. Finally, the mesh element identifier is stored, which forms a basis together with the photon barycentric coordinates for re-using a given photon for neighbouring frames. We assume that the motion of objects between the subsequent frames is small enough that even the photons which are “attached” to moving mesh elements approximate the indirect lighting within these elements well. In Section 5.1 we present our solutions that prevent re-using of photons for neighbouring frames when the variations in indirect lighting are significant enough to be perceived by the observer.

Since mesh elements are the framework for our temporal processing, the photons are sorted according to the elements which they hit. Apart from storing photon records, an additional table which summarizes the number of photons per element is also kept for every frame. Those tables are later used to efficiently derive some statistics on lighting variations in the temporal domain for a given element (refer to Section 5.1). With the help of these statistics we can detect significant changes of lighting. In this case reusing photons for reconstructing the lighting in adjacent frames might be restricted. Using the tables for the purpose of such statistics means that effectively the exact positions of photon hit points within an element are ignored (only the photon number counts), however, the processing of complete photon records can be avoided. In the following section we describe those statistics in more detail. We also explain the spatiotemporal processing of photons for dynamic environments.

5 Spatiotemporal Photon Processing

In our technique we assume that photons are traced sparsely for all animation frames and our goal is to minimize the number of those photons without compromising the animation quality. To achieve this goal we exploit the temporal coherence of indirect lighting and for a given frame we also consider photons that were traced for neighboring frames. Ideally, as many frames should be processed as it is required to reduce the stochastic noise below the sensitivity level of the human observer. However, the expansion of the photon collection in the temporal domain might be limited due to changes in dynamic environments that affect the lighting distribution. A contradictory requirement arises between maximizing the number of collected photons and minimizing the number of neighbouring frames (the time span) for which these photons were traced. A trade-off solution to this problem relies on balancing the stochastic noise (resulting from collecting too few photons) and the errors in reconstructed illumination (caused by collecting too many invalid photons in the temporal domain) to make those artifacts as little objectionable as possible for the human observer. The perception-based AQM is used to find the minimal number of photons per

frame which is required to make the noise undetectable. An energy-based stochastic error metric, which is applied to each mesh element and to every frame, is used to guide the photon collection in the temporal domain. We found this mesh-element level of applying the energy-based metric to be very efficient, and therefore abandoned the use of perception-based guidance of photon collection at this low level which would be far more expensive.

We describe our energy-based error metric which controls the temporal photon processing and reduces the probability of using invalid photons in the scene regions in which lighting changes rapidly in Section 5.1. In Section 5.2 we discuss our techniques of spatiotemporal photon processing, which guarantee that the quality of the indirect lighting reconstruction is consistent through the whole animation.

5.1 Error Metric for Temporal Processing

As we argued in the previous section collecting photons in the temporal domain makes sense only if the lighting distribution does not change too rapidly for subsequent frames. We attempt to detect such changes locally on the level of single mesh elements. The practical question how to distinguish the actual changes in lighting from the stochastic error arises. This problem is especially difficult in our technique because we compute a very small number of photons for every frame, which results in high levels of noise. In practice, this means that only lighting changes that are significantly higher than the noise level can be detected, which requires estimating the noise.

If we assume for a moment that the lighting does not change between subsequent frames, then hitting mesh elements by photons can be modelled well by the Poisson distribution [2]. Since the mesh elements are small the probability p of hitting a given mesh element by a photon is also small, i.e., $p \ll 1$ as required by the Poisson process. Also, different photons hitting a mesh element are mutually independent, i.e., the probability of the same photon hitting a mesh element again as a result of its multiple reflections is small. The Poisson distribution only has a single parameter, the mean μ , which can be estimated as the mean number of photons hitting a mesh element. The standard deviation σ can simply be derived as $\sigma = \sqrt{\mu}$. Thus, the noise level can be estimated as $\mu \pm k\sigma$, where e.g., $k = 2$ (for $\mu = 0$ we assign $\sigma = 1$). Based on this estimate we assume that if the number of photons x hitting a mesh element does not satisfy the condition

$$\mu - k\sigma \leq x \leq \mu + k\sigma \quad (1)$$

a change of lighting can be expected and the photon collection for this mesh element is disabled. For a given mesh element the mean μ is estimated for the currently considered frame and the values of x are obtained for the corresponding mesh element in the preceding and following frames. The temporal collection of photons is initiated from the current frame and proceeds for subsequent frames as long as Condition (1) is satisfied. The photon processing is performed independently in the directions of the preceding and following frames. Thus, the photon collection for a given mesh element may be asymmetric with respect to the current frame when Condition (1) is violated earlier in one of those directions than in another one.

There are many possible sources of error, which may prevent the application of Condition (1) from working robustly for all mesh elements. First of all the estimate of μ might be inexact since we use a Monte Carlo solution at very early stages of convergence. To reduce the influence of outliers on the estimate of μ we consider the mean number of photons hitting a given mesh element for three subsequent frames. We apply this procedure to derive x for the same reason. More than three subsequent frames might be considered but then the μ estimate might be affected by changes in lighting.

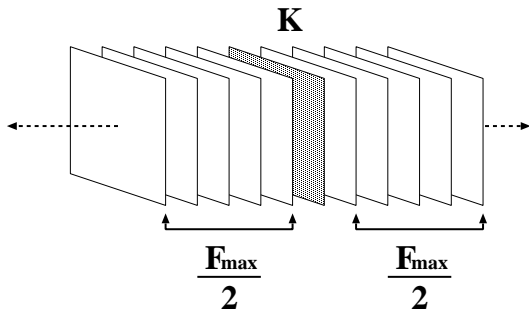


Figure 2: Organization of an animation segment.

Clearly, the procedure works the better the more photons per frame are computed.

If the average number of photons per mesh element is small, assigning $k > 2$ might be considered to compensate for inaccuracies in the μ estimates. Also, small lighting changes on the level $\pm k\sigma$ are usually less perceptible than the temporal aliasing which becomes visible when the temporal photon collection is prematurely disabled. In practice, we assume the minimal extent of the temporal processing to be at least 3 frames. Even if the lighting conditions change drastically for a pair of subsequent frames, the visibility thresholds are usually elevated due to the temporal masking for up to 100 milliseconds [10]. Since new lighting details cannot be seen well for 2–3 animation frames the lighting reconstruction accuracy can be relaxed for those frames, and tracing additional photons can be avoided.

In the following section we describe the algorithm of animation rendering, which extensively uses the procedure of adaptive photon collection in the temporal domain described in this section.

5.2 Algorithm

The animation sequence is split into animation segments, which are then processed sequentially one by one. The organization of a single segment is depicted in Figure 2. The frame K divides the segment into two halves of length $F_{\max}/2$ each. The goal of the segment processing is to reconstruct indirect lighting for all $F_{\max} + 1$ frames. To enable temporal processing of the whole segment the photons collected for neighbouring segments for up to F_{\max} frames preceding K and up to F_{\max} frames following K are also considered. For the very first and last animation segments the lack of those extra frames is compensated by tracing more photons.

The processing of every animation segment consists of five major steps:

1. Initialization: determination of the number of initial photons per frame, and processing of those photons for all frames in the current segment.
2. Adjustment of the segment length depending on the temporal variations of indirect lighting which are measured using energy-based criteria (refer to Section 5.1).
3. Adjustment of the number of photons per frame based on the AQM response in order to limit the perceivable stochastic noise to an acceptable level.
4. Spatiotemporal reconstruction of indirect lighting for all mesh elements guided by the same energy-based criteria as in Step 2.
5. Spatial filtering step for those mesh elements that did not meet the perceptual and energy criteria in the previous two steps.

In practice, Steps 1–3 are performed not only for the current segment but for the subsequent segment as well, which results in the processing of all photons that are used for indirect lighting reconstruction in the current segment at Step 4. Obviously, photons for the preceding segment that has already been processed are also available for the temporal processing of frames in the current segment.

In the following sections we describe all steps in detail.

5.2.1 Initialization

In the initialization step we decide how many photons N_{frame} per frame have to be shot for a given animation segment. Also, the initial segment length F_{\max} needs to be chosen based on the anticipated complexity of the lighting changes during an animation. The initial settings of F_{\max} and N_{frame} are adjusted later for every animation segment using the energy-based and perception-based error metrics. However, the better the initial guess is, the smaller number of iterations involving those metrics is needed. In our approach F_{\max} is set manually by the user for the first processed segment. The adjusted F_{\max} for a previous segment is assigned as the initial length of a subsequent segment. Based on F_{\max} chosen for a given segment, N_{frame} is decided automatically as follows.

An initial value for N_{frame} should take into account basic lighting characteristics of the considered animation segment. To derive such an initial guess we applied the energy-based measure of the error of indirect lighting simulation¹ E . We observed that by setting the maximum error value E_{\max} to lie inside the range of 1–5% we usually obtained images of good quality. Further computation basically did not introduce any improvement as perceived by the human observer [27]. In practice, we assume a less conservative $E_{\max} = 5\%$. Now, if we would run the DEPT computation until $E \approx E_{\max}$ we would find the number of photons N_{\max} that is required to achieve this accuracy. Since we want to reconstruct indirect lighting for a given frame using photons computed for up to $F_{\max}/2$ preceding and $F_{\max}/2$ following frames, we could estimate $N_{\text{frame}} = N_{\max}/(F_{\max} + 1)$. In practice, we do not want to run the computation for a single frame long enough to reach the error level E_{\max} . However, we can directly get a good estimate of N_{frame} much faster by using the basic property of stochastic solution convergence stating that the error is proportional to the inverse square root of the number of traced photons [15]:

$$\frac{E}{E_{\max}} \approx \sqrt{\frac{N_{\max}}{N_{\text{frame}}}} \quad (2)$$

This means that N_{frame} can be estimated by running the pilot DEPT computation until $E \approx E_{\max}\sqrt{F_{\max} + 1}$.

Finally, for all frames in the current animation segment N_{frame} photons are traced, and their hit point records are sorted for every mesh element.

5.2.2 Choosing the Animation Segment Length

The main goal of adjusting the maximum segment length F_{\max} is to reduce the animation artifacts caused by collecting photons in the temporal domain over too many frames, as this may lead to an invalid lighting reconstruction. For the scene regions in which the temporal changes of the lighting function are fast, the collection of photons can be performed only for a small number of subsequent frames. Conversely, for the scene regions in which the temporal changes are slow the photons should be collected for as many frames as possible to reduce the stochastic noise.

¹The detailed description of the stochastic method used to estimate the lighting simulation error in the framework of the DEPT algorithm can be found in [28].

In Section 5.1 we described our procedure of adaptive photon collection in the temporal domain. The maximum number of frames for which photons are collected using this procedure is limited by the segment length F_{\max} . Increasing F_{\max} incurs additional costs of processing a larger number of frames, and is justified only when collecting photons from F_{\max} frames is possible for a significant percentage p_{me} of the mesh elements. In practice, the user sets a certain threshold value p_{thr} , e.g., 40%, and F_{\max} is increased until the condition $p_{\text{me}} < p_{\text{thr}}$ is met. Such a test is performed only once per segment, when the lighting is reconstructed for the central keyframe K . Note that increasing F_{\max} causes moving K apart $F_{\max}/2$ frames from the end of the previously processed segment (or from the beginning of animation sequence for the very first processed segment). The F_{\max} value obtained for K is then used for all frames in the current segment.

5.2.3 Choosing the Number of Photons

The main goal of adjusting the number of photons per frame N_{frame} is to keep the noise below a perceivable level. Increasing N_{frame} for each frame in a given segment is an expensive operation and should be performed if the adjustment of F_{\max} performed in the previous step did not provide the required animation quality as perceived by the human observer.

The AQM is used to measure the errors of the indirect lighting reconstruction which enables the perception-guided selection of N_{frame} to minimize the computational costs without degrading the animation quality. The AQM requires two animation frames I_1 and I_2 as input, and will then predict the perceivable differences between them. Ideally, a frame resulting from the temporal photon processing should be compared to some exact reference solution. Since such a reference solution is not available in practice, we decided to measure the differences in indirect lighting reconstructed for the central frame K by splitting the photons collected for all frames in a given segment into two halves (the even and odd photons). The indirect lighting in $I_1(K)$ and $I_2(K)$ is reconstructed using these halved sets of photons. In order to measure the level of noise in the conditions in which the actual temporal photon processing is performed for all animation frames, the procedure of adaptive photon collection in the temporal domain is used for the $I_1(K)$ and $I_2(K)$ computation as well (refer to Section 5.1).

The approach of halving sets of photons is quite conservative because according to the Monte Carlo methods theory [15] the stochastic error of indirect lighting reconstruction in the frame $I(K)$ that is obtained for 100% of photons is smaller by the factor $\sqrt{2}$ with respect to $I_1(K)$ and $I_2(K)$. The perceivable differences as predicted by the AQM usually are reduced by an even larger factor if the number of photons is doubled.

As the result of AQM processing a map M_{AQM} is generated which shows the prediction of perceivable differences in indirect lighting between all corresponding pixels in $I_1(K)$ and $I_2(K)$. As a scalar metric of the frame quality the percentage d_{AQM} of M_{AQM} pixels with differences over one unit Just Noticeable Difference (JND) is assumed [18]. The user chooses a certain threshold value d_{thr} of the AQM predicted differences, and when $d_{\text{AQM}} > d_{\text{thr}}$, N_{frame} is doubled and the whole procedure is repeated until $d_{\text{AQM}} < d_{\text{thr}}$.

To reduce the costs of Human Visual System (HVS) modelling the AQM processing is performed only once per segment for the central frame K . Thus, the N_{frame} value obtained for K is assumed to be valid for all frames in a given segment. In practice, this trade-off approach works well because the differences in indirect lighting are usually small for a given animation segment whose length F_{\max} was adjusted to reduce such differences (refer to Section 5.2.2).

5.2.4 Indirect Lighting Reconstruction

After establishing F_{\max} and N_{frame} further processing of all frames in a given segment becomes straightforward. At first, the valid photons are collected for each mesh element, using the procedure of adaptive photon collection in the temporal domain described in Section 5.1. Then the standard procedure for deriving illumination at mesh vertices is applied (refer to Section 4).

5.2.5 Repairing Noisy Pixels

The procedure described in the previous section may potentially result in locally noisy images² for the scene regions (e.g., moving objects) in which the indirect lighting changes much faster than for the remaining parts of environment. For such regions collecting photons in the temporal domain is usually limited to a few subsequent frames. Obviously, the noise level could be reduced by increasing N_{frame} which is costly. Note that in the procedure of selecting N_{frame} using the AQM we allow perceivable differences for up to d_{thr} pixels (refer to Section 5.2.3). When the perceivable problems concern only a small fraction of mesh elements, then for efficiency reasons, increasing N_{frame} for all frames in a given segment should be avoided.

Our solution relies on using the spatial filtering performed in the object space selectively for those mesh elements for which the expansion in the temporal domain was not possible, resulting in a small number of collected photons. We apply the filtering algorithm which was originally proposed in [27]. To achieve the required level of accuracy of reconstructed lighting at a given vertex, photons hitting a region h centered at this vertex are considered. Stochastic variance estimates of the local illumination are used to decide upon the size of h . This effectively reduces noise, however some bias is introduced to the reconstructed lighting (refer to [27] for a formal derivation of a mathematically-sound measure of illumination accuracy and a detailed description of the filtering algorithm).

Using spatial filtering is equivalent to trading in the spatial details of indirect lighting in order to remove excessive noise. If this approach is not acceptable, a final gathering step [14] could be performed. However, we did not apply this solution because of its significant cost. We found that the spatial filtering approach as applied in the indirect lighting reconstruction produces good results in terms of the animation quality as perceived by the human observer.

5.3 Accuracy Considerations

The accuracy of the indirect lighting reconstruction using our algorithm described in Section 5.2 is limited by the spatial resolution of the mesh used for collecting photons. Consequently the solution is biased. On the other hand the mesh resolution can be set arbitrarily fine and more photons can be traced. For those mesh elements that still collect too few photons in the temporal domain, the spatial filtering discussed in Section 5.2.5 can be used to remove visible noise. Another source of the bias is the temporal blur resulting from the collection of invalid photons in the temporal domain. The level of blurring can be controlled in the energy terms and traded for the stochastic noise by decreasing the value of parameter k in Condition (1).

The AQM is used to measure the perceivable differences between two equally biased indirect lighting solutions, which means that all measured differences between frames $I_1(K)$ and $I_2(K)$ result from the stochastic noise (refer to Section 5.2.3). Effectively the AQM provides a conservative stopping condition for photon tracing when the noise falls below the sensitivity level of the human observer. Tracing more photons cannot improve the perceived

²As a matter of fact for a vast majority of tests that we performed, we were not able to notice such problems.

quality of the indirect lighting reconstruction due to limitations in the spatial mesh resolution.

6 Results

We present results that we obtained for the ROOM scene (about 5,300 mesh elements). Also, we briefly summarize the results obtained for another scene ATRIUM (about 45,000 mesh elements), which are qualitatively very similar and therefore do not need to be discussed in full length. Both scenes were designed in such a way that moving objects significantly affected the lighting distribution. Also, some scene regions are illuminated exclusively by indirect lighting which imposes higher requirements on its reconstruction. We begin with discussing the experimental results for the adaptive algorithm of temporal photon processing discussed in Section 5.1. Then we discuss the step by step results obtained for the spatiotemporal photon processing discussed in Section 5.2.

The simplest scenario of temporal photon processing is to consider the fixed number of preceding and following frames. However, this approach may lead to significant errors as illustrated in Figures 3. Figure 3a shows the correct reference frame obtained for the converged DEPT solution. In this scene a spot light illuminates the bottom of the aircraft, and the highlight on the floor is caused by the light reflected from the aircraft. Note that as the result of non-adaptive temporal processing for all $F_{\max} = 30$ frames the highlight is significantly washed out (Figure 3b). When applying our adaptive photon collection technique (Section 5.1) the highlight shown in Figure 3c is similar to the reference frame in Figure 3a. Figure 3d shows the AQM produced map M_{AQM} , in which red color marks pixels for which visible differences are predicted. Figures 4a and b summarize the AQM predicted percentage of pixels d_{AQM} with perceivable differences derived from M_{AQM} for various settings of N_{frame} and F_{\max} for non-adaptive and adaptive photon collection approaches. As can be seen in Figure 4a for the non-adaptive approach, expanding F_{\max} initially leads to reducing d_{AQM} , but then the collection of invalid photons results in increasing d_{AQM} for large F_{\max} . The corresponding characteristics for the adaptive approach shown in Figure 4b are extremely favorable because the expansion of F_{\max} always leads to the reduction of d_{AQM} , which means that collecting invalid photons is mostly avoided.

Following the subsequent processing steps described in Section 5.2 we obtained the following animation settings for the ROOM scene. As the result of the initialization procedure $N_{\text{frame}} = 10,000$ and $F_{\max} = 15$ were chosen. The animation was split into three segments and the final settings computed for the central segment frames K_i are summarized in Table 1. In segment K_3 N_{frame} is smaller and F_{\max} more expanded because strong direct lighting washes out some imperfections of the indirect lighting reconstruction. Also, changes of lighting are less dynamic in this animation segment.

Figure 5 summarizes the results obtained for K_2 using the energy-based procedure of photon validity estimation for subsequent frames described in Section 5.2.2. We assumed $p_{\text{thr}} = 40\%$. The maximum segment length $F_{\max} = 30$ was chosen (refer to Table 1), in which case $p_{\text{me}} = 36\%$ and $p_{\text{me}} = 30\%$ were obtained for the preceding and following directions, respectively.

	N_{frame}	F_{\max}
K_1	40,000	30
K_2	40,000	30
K_3	10,000	44

Table 1: Final settings for the ROOM scene animation.

Figure 6 summarizes the results obtained using the perception-based procedure of noise level estimation as described in Section 5.2.3. It was assumed that $d_{\text{thr}} = 3\%$, which means in practice that the perceivable differences $d_{\text{AQM}} < 1\%$ with respect to the reference solution are usually obtained. Table 1 summarizes the number of photons N_{frame} chosen for every segment based on the graphs in Figure 6. For such animation settings the spatial filtering described in Section 5.2.5 was not necessary.

Figure 7a shows an animation frame obtained using the settings presented in Table 1. Figure 7b depicts the corresponding frame obtained using the traditional approach without any temporal photon processing. Strong temporal aliasing was observed when the animation composed of such quality frames was viewed. We also tried the traditional approach with $N_{\text{frame}} = 845,000$ which corresponds to the average number of photons collected in the temporal domain using our approach. While the static image is of a quality comparable to the frame in Figure 7a some temporal aliasing can be seen clearly when the resulting animation is viewed.

The results obtained for the ATRIUM scene are very similar to the ones for ROOM. For a majority of segments $N_{\text{frame}} = 20,000$ and $F_{\max} = 44$ were chosen using our automatic procedures described in Section 5.2. In general, N_{frame} fell into the range of 10,000–40,000 photons while F_{\max} lay between 30 and 44 frames. The activation of spatial filtering (Section 5.2.5) led to some minor improvement of the animation quality. Figure 8 shows an example of a frame with spatial filtering and the only differences with respect to the corresponding frame without filtering can be seen on the stairs which feature small mesh elements and little visual masking. Refer to our Web site www.mpi-sb.mpg.de/resources/aqm/dynenv/ for the samples of all animations discussed in this section.

A summary of timings of indirect lighting computation is given in Table 2. As can be seen, tracing photons and their temporal processing is rather inexpensive. The higher cost of temporal processing for the ATRIUM scene is due to the larger number of processed mesh elements. The I/O costs related to disk access are given for two extreme scenarios denoted in Table 2 as MIN and MAX (provided in brackets). In the former case photons are accessed from the disk only once and then cached until all frames using those photons have been processed. This means that all photons processed for a given frame must reside in memory, which is a common requirement for many photon-based techniques such as the photon map [14]. This is usually a reasonable assumption for our technique as well. For example, the maximum memory storage per frame for the ATRIUM sequence was about 120 MB (one photon hit point requires 20 bytes of storage). In the case denoted by (MAX) in Table 2 it was assumed that all photons are always loaded from disk for each frame. Such a situation may arise for complex scenes when a high accuracy of lighting reconstruction is required, in which case a large number of photons that are bucketed into a fine mesh must be considered (refer to Section 5.3). Note that even in such a case our timings are at least three times better than shooting a similar number of photons that we collected in the temporal domain for every frame, which requires 87 and 133 seconds for the ROOM and ATRIUM scenes, respectively. Ray tracing of a single frame requires 9.4 seconds for ROOM and 158 seconds for ATRIUM.

Scene	Photon tracing	AQM	Temp. proc.	I/O	Total MIN
				MIN (MAX)	
ROOM	2.57	0.27	0.32	0.72 (21.56)	3.88
ATRIUM	2.95	0.21	1.85	0.88 (26.93)	5.89

Table 2: Timings of the indirect lighting computation for a single frame obtained as the average cost per frame for the whole animation. All timings are given in seconds and were measured on a 800 MHz Pentium III processor.

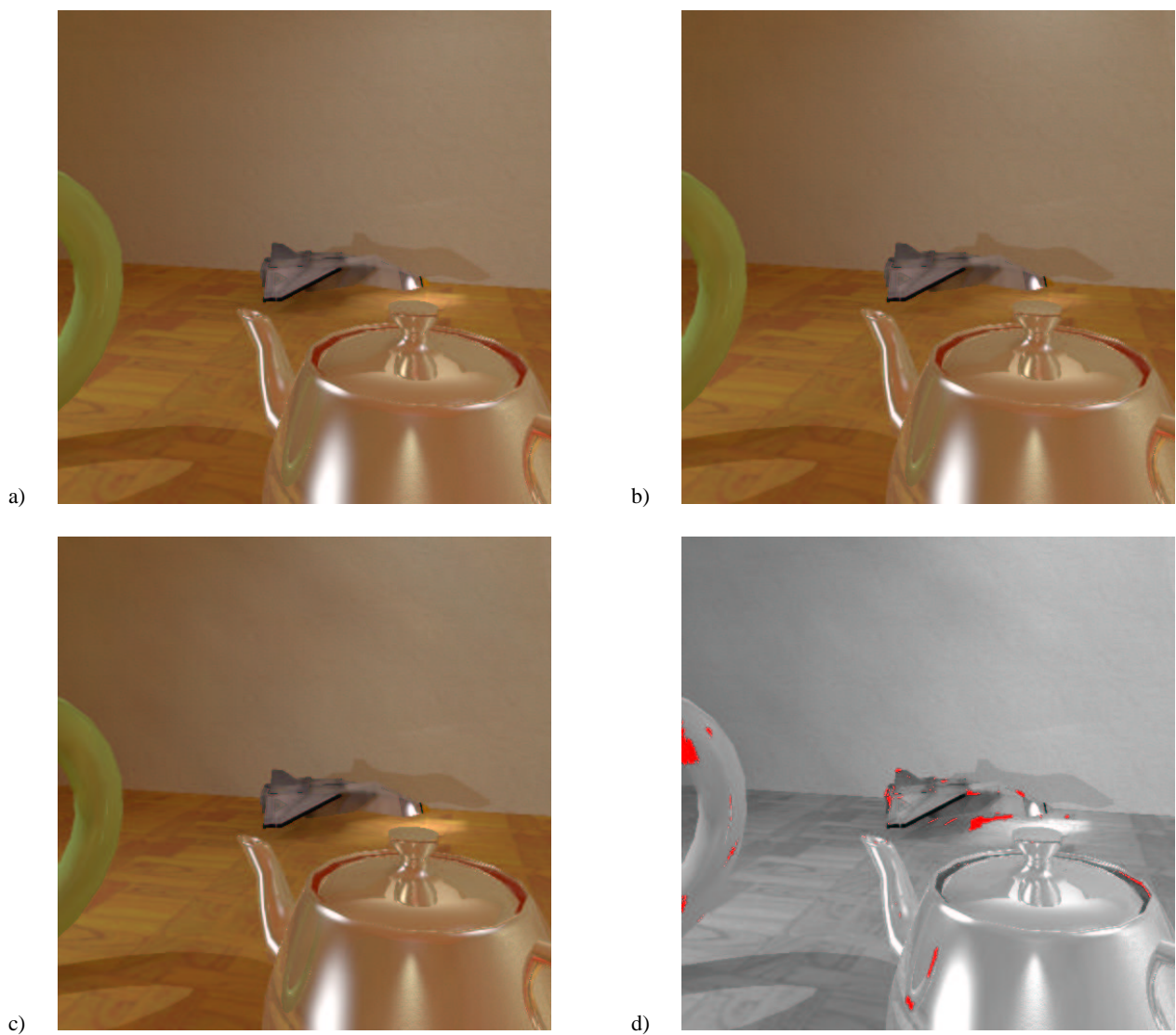


Figure 3: Example frame from the ROOM sequence a) reference solution for 2,000,000 photons without temporal processing, b) non-adaptive (note the washed-out highlight under the plane) and c) adaptive photon collection in the temporal domain for $F_{\max} = 30$ frames and $N_{\text{frame}} = 40,000$ photons, and d) the map of AQM predicted perceivable differences (marked in red) between a) and c).

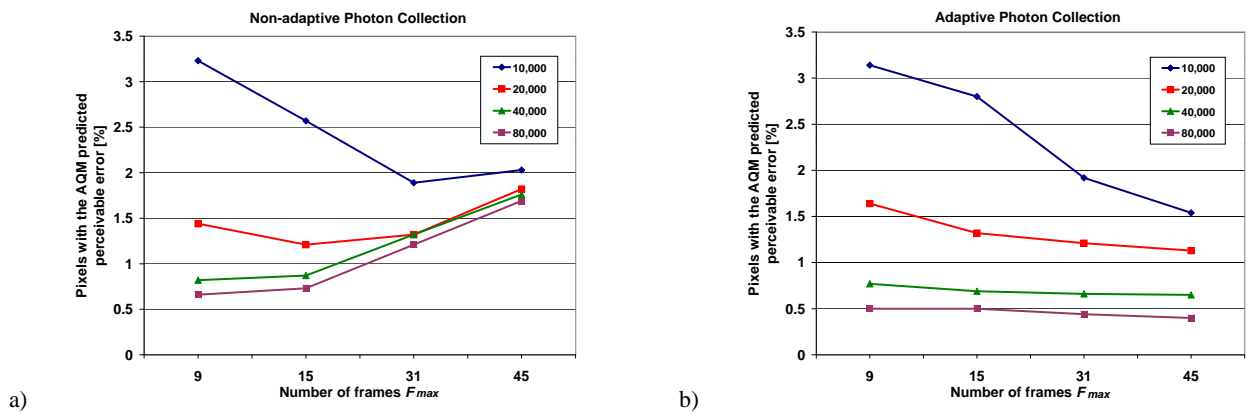


Figure 4: The AQM predicted percentage of pixels d_{AQM} with perceivable differences for a) non-adaptive and b) adaptive temporal photon collection approaches for increasing F_{\max} and various settings of N_{frame} (as specified in the legend).

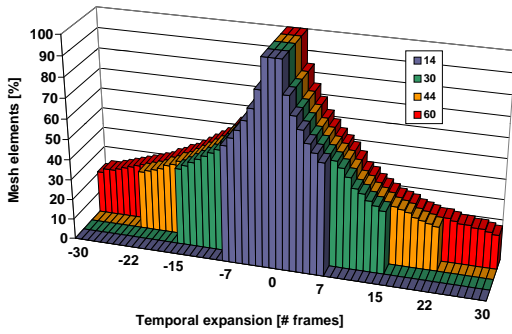


Figure 5: Distribution of mesh elements for frame K_2 as a function of the number of preceding (negative values) and following frames for which temporal photon processing was possible. Various F_{\max} were considered as specified in the legend.

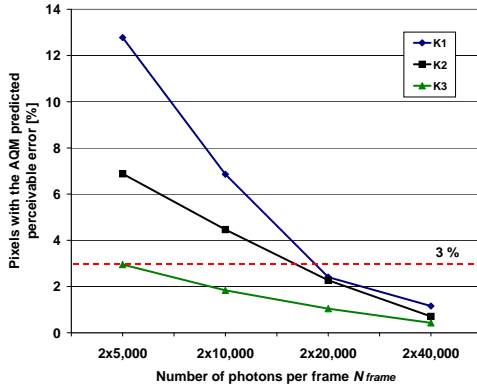


Figure 6: The AQM predicted percentage of pixels d_{AQM} with perceivable differences as a function of N_{frame} for the central segment frames K_i .

Another important issue is the cost of perception-based error metrics. The choice of characteristics of the HVS used to guide lighting computations should be weighted by the possible performance gains obtained by their consideration. For example, timings reported by Yee for rendering a single frame using the RADIANCE system including indirect lighting computation amounted to 3–6 hours, which means that the overhead of the HVS modelling is marginal. Our indirect lighting solution requires only single minutes for similar scenes so that the overhead becomes significant. In practice, this means that the HVS characteristics that are too costly to model have to be ignored. In order to reduce the cost of the AQM, we ignore the orientation channels processing in the visual masking, which to some extent can be justified by cross-masking between various orientations as discussed in [20]. Also, we scale down images which are input to the AQM to 256×256 pixels. At the same time we proportionally reduce the distance of the observer to the screen (which is an input parameter of the AQM) to preserve the same spatial frequencies as in the original animation frames. We compute the pixel flow, which is needed to model the spatio-velocity Contrast Sensitivity Function (an important component of the AQM), by compensating the camera motion using 3D warping of the frames neighboring to a given keyframe as in [18]. We assume that the motion of animated objects is fully compensated by the smooth pursuit eye motion (thus, the computation of pixel flow is not required, and the significant cost of such a computation [31] can be avoided), which leads to the high sensitivity of the eye for such objects. This assumption is justified by the fact that moving objects

are one of the most important attractors of the visual attention [20], which means that efficiency of the eye tracking for such objects is very high. All those simplifications result in the AQM processing time of about 4 seconds for a pair of compared frames.

7 Conclusions

We proposed a novel global illumination technique for dynamic environments which is suitable for high-quality animation rendering. A combination of efficient energy- and perception-based error metrics was used to guide the computation as a function of local spatiotemporal variations of the lighting distribution. As a result the animation quality as perceived by a human observer is consistent across all frames both in spatial and temporal dimensions. Also, the efficiency of computation is improved and the temporal aliasing is reduced with respect to traditional approaches which ignore temporal processing.

As future work we want to investigate our technique in the context of MPEG coding. The accuracy of the lighting reconstruction can be adjusted in order to obtain a degradation of the animation quality that is perceived as being as homogeneous as possible for an assumed animation compression level. Also, by removing non-visible lighting details from animations the compression performance can be improved.

8 Acknowledgments

We would like to thank Kirill Dmitriev and Edward Kopylov for help in interfacing our software to the Inspirer system, Integra Inc., which we used as a testbed in this work. Thanks also to Philippe Bekaert, Katja Daubert, Annette Scheel, and Hartmut Schirmacher for their helpful comments and suggestions. This work was supported in part by the ESPRIT Open LTR Project #35772 SIMUL-GEN “Simulation of Light for General Environments”.

References

- [1] A.A. Apodaca and L. Gritz. *Advanced RenderMan*. Morgan Kaufmann, 1999.
- [2] P.R. Bevington and D.K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, Inc., New York, 1992.
- [3] M.R. Bolin and G.W. Meyer. A Perceptually Based Adaptive Sampling Algorithm. In *SIGGRAPH 98 Conference Proceedings*, Annual Conference Series, pages 299–310, 1998.
- [4] S.E. Chen. Incremental Radiosity: An Extension of Progressive Radiosity to an Interactive Image Synthesis System. In *Computer Graphics (SIGGRAPH 90 Conference Proceedings)*, pages 135–144, 1990.
- [5] C. Damez and F.X. Sillion. Space-Time Hierarchical Radiosity for High-Quality Animations. In *Eurographics Rendering Workshop 1999*, pages 235–246, 1999.
- [6] G. Drettakis and F.X. Sillion. Interactive Update of Global Illumination Using a Line-Space Hierarchy. In *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 57–64, 1997.
- [7] J.A. Ferwerda, S. Pattanaik, P. Shirley, and D.P. Greenberg. A Model of Visual Masking for Computer Graphics. In *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 143–152, 1997.
- [8] D.W. George, F.X. Sillion, and D.P. Greenberg. Radiosity Redistribution for Dynamic Environments. *IEEE Computer Graphics and Applications*, 10(4):26–34, July 1990.
- [9] S. Gibson and R.J. Hubbard. Efficient Hierarchical Refinement and Clustering for Radiosity in Complex Environments. *Computer Graphics Forum*, 15(5):297–310, 1996.
- [10] B. Girod. The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals. pages 178–187. Proc. of SPIE Vol. 1077, 1989.
- [11] D.P. Greenberg. A Framework for Realistic Image Synthesis. *Communications of the ACM*, 42(8):43–53, August 1999.

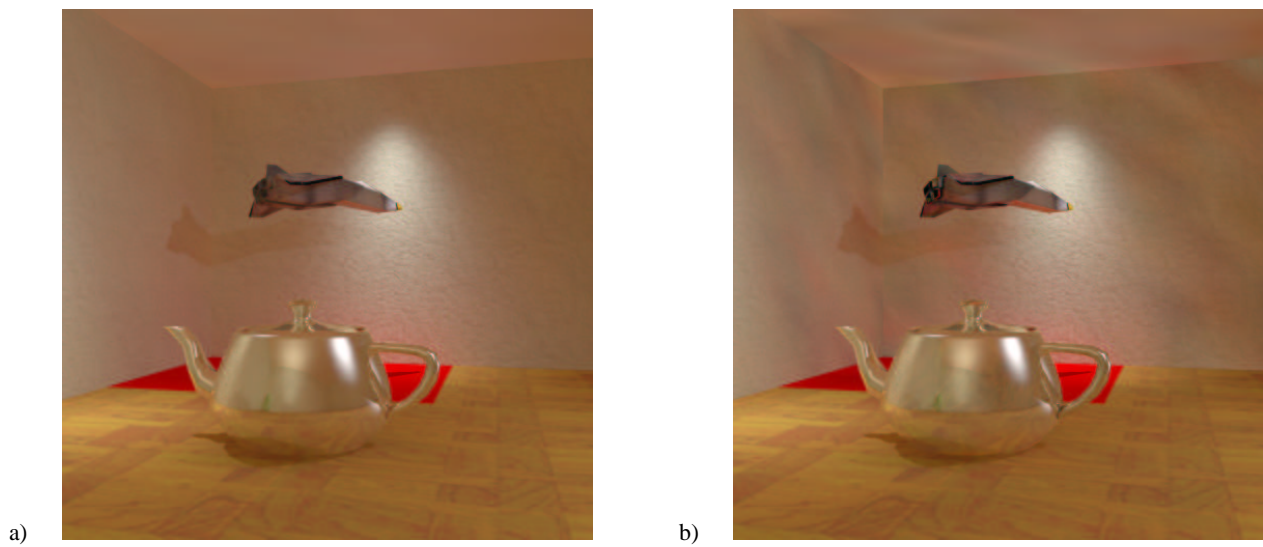


Figure 7: Example frame from the ROOM sequence a) with temporal processing for $N_{\text{frame}} = 10,000$ and $F_{\text{max}} = 44$ and b) without temporal processing for $N_{\text{frame}} = 24,550$. The same computation time was spent to generate those frames, but in b) the average overhead time of 1.31 seconds, which was required in a) for the temporal, AQM, and I/O processing (refer to Table 2), was used to trace more photons.

[12] P. Heckbert. Adaptive Radiosity Textures for Bidirectional Ray Tracing. In *Computer Graphics (SIGGRAPH 90 Conference Proceedings)*, pages 145–154, August 1990.

[13] W. Heidrich. Interactive Display of Global Illumination Solutions for Non-Diffuse Environments. In *State of the Art Reports*. Eurographics, 2000.

[14] H. W. Jensen. Global Illumination Using Photon Maps. In *Eurographics Rendering Workshop 1996*, pages 21–30, 1996.

[15] M.M. Kalos and P.A. Whitlock. *Monte Carlo Methods*. Wiley International, 1986.

[16] A. Keller. Instant Radiosity. In *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 49–56, 1997.

[17] S. Müller, W. Kresse, and F. Schoeffel. A Radiosity Approach for the Simulation of Daylight. In *Eurographics Rendering Workshop 1995*, pages 137–146, 1995.

[18] K. Myszkowski, P. Rokita, and T. Tawara. Perceptually-Informed Accelerated Rendering of High Quality Walkthrough Sequences. In *Eurographics Rendering Workshop 1999*, pages 5–18, 1999.

[19] J. Nimeroff, J. Dorsey, and H. Rushmeier. Implementation and Analysis of an Image-Based Global Illumination Framework for Animated Environments. *IEEE Transactions on Visualization and Computer Graphics*, 2(4):283–298, 1996.

[20] W. Osberger. *Perceptual Vision Models for Picture Quality Assessment and Compression Applications*. Ph.D. thesis, Queensland University of Technology, 1999.

[21] X. Pueyo, D. Tost, I. Martin, and B. Garcia. Radiosity for Dynamic Environments. *The Journal of Visualization and Comp. Animation*, 8(4):221–231, 1997.

[22] M. Ramasubramanian, S.N. Pattanaik, and D.P. Greenberg. A Perceptually Based Physical Error Metric for Realistic Image Synthesis. In *SIGGRAPH 99 Conference Proceedings*, Annual Conference Series, pages 73–82, 1999.

[23] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing Real and Synthetic Images: Some Ideas About Metrics. In *Eurographics Rendering Workshop 1995*, pages 82–91, 1995.

[24] F. Schoeffel and P. Pomi. Reducing Memory Requirements for Interactive Radiosity Using Movement Prediction. In *Eurographics Rendering Workshop 1999*, pages 225–234, 1999.

[25] P. Shirley, B. Wade, P. M. Hubbard, D. Zareski, B. Walter, and D. P. Greenberg. Global Illumination via Density Estimation. In *Eurographics Rendering Workshop 1995*, pages 219–230, 1995.

[26] J. Tumblin and H.E. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993.

[27] V. Volevich, K. Myszkowski, A. Khodulev, and Kopylov E.A. Using the Visible Differences Predictor to Improve Performance of Progressive Global Illumination Computations. *ACM Transactions on Graphics*, 19(2):122–161, 2000.

[28] V. Volevich, K. Myszkowski, A.B. Khodulev, and E.A. Kopylov. Perceptually-Informed Progressive Global Illumination Solution. Technical Report TR-99-1-002, Department of Computer Science, Aizu University, February 1999.

[29] B.J. Walter. *Density Estimation Techniques for Global Illumination*. Ph.D. thesis, Cornell University, 1998.

[30] G.J. Ward. The RADIANCE Lighting Simulation and Rendering System. In *SIGGRAPH 94 Conference Proceedings*, Annual Conference Series, pages 459–472, 1994.

[31] Y.L.H. Yee. *Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments*. M.Sc. thesis, Cornell University, 2000.

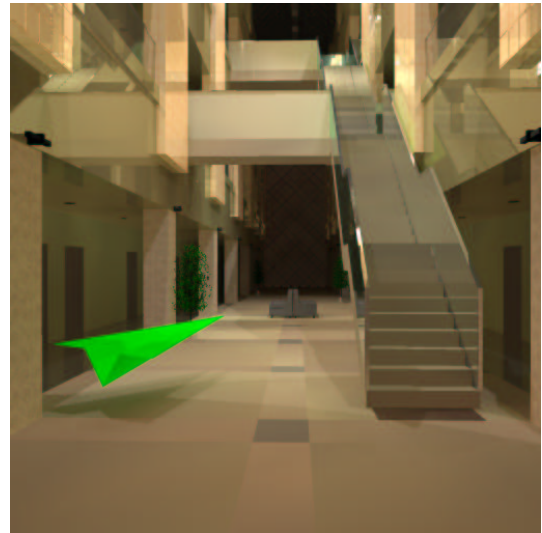


Figure 8: Example frame from the ATRIUM sequence with temporal processing and spatial filtering for $N_{\text{frame}} = 20,000$ and $F_{\text{max}} = 44$.