

# Perception-Based Global Illumination, Rendering, and Animation Techniques

Karol Myszkowski

Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany  
karol@mpi-sb.mpg.de

## Abstract

In this paper, we consider applications of perception-based video quality metrics to improve the performance of global lighting computations and rendering of animation sequences. To control the computation of animation frames we use the Animation Quality Metric (AQM) which is an extended version of the Visible Difference Predictor (VDP) developed by Daly. We show two applications of the AQM: (1) the rendering of high-quality walkthroughs for static environments and (2) the computation of global illumination for dynamic environments.

To improve the rendering performance of our walkthrough solution we use a hybrid of the ray tracing and Image-Based Rendering (IBR) techniques. In our rendering solution we derive as many pixels as possible using inexpensive IBR techniques without affecting the animation quality. The AQM is used to automatically guide such a hybrid rendering.

Also, 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. The AQM 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.

**Keywords:** Video quality metrics, global illumination, realistic rendering, temporal processing

## 1 Introduction

Rendering of animated sequences proves to be a very computation intensive task. In professional production this involves specialized rendering farms designed specifically for this purpose. Data revealed by major animation companies show that rendering times for the final antialiased frames are still counted in tens of minutes or hours [1], so shortening this time becomes very important. A serious drawback of traditional approaches to animation rendering is that error metrics controlling the quality of frames

(which are computed separately one by one) are too conservative, and do not take advantage of various limitations of the Human Visual System (HVS).

It is well-known in the video community that the human eye is less sensitive to higher spatial frequencies than to lower frequencies, and this knowledge was used in designing video equipment [30]. It is also conventional wisdom that the requirements imposed on the quality of still images must be higher than for images used in an animated sequence. Another intuitive point is that the quality of rendering can usually be relaxed as the velocity of the moving object (visual pattern) increases. These observations are confirmed by systematic psychophysical experiments investigating the sensitivity of the human eye for various spatiotemporal patterns [12, 34]. This means that all techniques attempting to speed up the rendering of every single frame separately cannot account for the eye sensitivity variations resulting from temporal considerations. Effectively, computational efforts can be easily wasted on processing image details which cannot be perceived in the animated sequence. In this context, a global approach involving both spatial and temporal dimensions appears promising research direction. In this paper we discuss two example solutions in which spatio-temporal considerations are taken into account to improve the performance of walkthrough animation rendering [21] and global illumination computation for dynamic environments [22].

In the following section we briefly discuss previous work on the perception-guided animation solutions and video quality metrics. Then we present our extensions of the VDP that are required to develop the AQM which is suitable to handle animated sequences. In Section 4 we show the AQM application to improve the performance of rendering animation walkthrough sequences. In Section 5 we present the AQM-based guidance of photon tracing which improves the efficiency of indirect lighting computations for high-quality animation sequences. Also, we discuss the obtained results using our techniques. Finally, we conclude the paper and propose some future directions for this research.

## 2 Previous Work

### 2.1 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. In recent years some successful examples of perception-based rendering of static images have been presented [8, 2, 25], however, expanding those techniques to handle the temporal domain remains mostly an open problem.

Recently, some attempts of using higher level perceptual and cognitive elements have been introduced in the context of animation. Yee et al. [38] propose an interesting application of a visual attention model to improve the efficiency of indirect lighting computations in the RADIANCE system [33] for dynamic environments. A saliency map based on the visual attention model developed by Itti [11] is used to control the caching of secondary lighting in the RADIANCE system on a per pixel basis. For image regions less salient, greater errors can be tolerated, and the indirect lighting can be interpolated for a larger neighborhood which makes caching more efficient at the expense of blurring details in the lighting distribution. Yee et al. demonstrated that by considering visual attention significant rendering speedup can be achieved. 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 et al. report 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.

### 2.2 Video Quality Metrics

Assessment of video quality in terms of artifacts visible to the human observer is becoming very important in various applications dealing with digital video encoding, transmission, and compression techniques. Subjective video quality measurement usually is costly and time-consuming, and requires many human viewers to obtain statistically meaningful results [29]. In recent years, a number of automatic video quality metrics, based on the computational models of human vision, has been proposed. Some of these metrics were designed for video [30, 37], and are often specifically tuned [39] for the assessment of perceptibility of typical distortions arising in lossy video compression such as blocking artifacts, blurring, color shifts, and fragmentation. Also, some well-established still im-

age quality metrics were extended into the time domain [15, 35, 29].

In this study, we deal exclusively with synthetic images, and we are looking for a metric well tuned to our application requirements, even at the expense of some loss of its generality. We found that such a metric is the VDP, a static image quality metric proposed by Daly [3], which we extend to handle animated sequences. In the following section we describe in detail our extensions of the VDP into the temporal domain.

## 3 Animation Quality Metric

Before we move on to the description of the VDP customizations (Section 3.3), we recall basic facts on the spatiotemporal Contrast Sensitivity Function (CSF) which is an important component of virtually all advanced video quality metrics. We show that in our application it is far more convenient to use the spatiovelocity CSF, which is a dual representation of the commonly used spatiotemporal CSF (Section 3.1). Also, we present our algorithm for efficient computation of the velocity of patterns moving across the image plane in animated sequences (Section 3.2).

### 3.1 Spatiovelocity CSF Model

Spatiotemporal sensitivity to contrast, which varies with the spatial and temporal frequencies is an important characteristic of the HVS. The sensitivity is characterized by the so called spatiotemporal CSF, which defines the detection threshold for a stimulus as a function of its spatial and temporal frequencies. The spatiotemporal CSF is widely used in multiple applications dealing with motion imagery. One of the most commonly used analytical approximations of the spatiotemporal CSF are the formulas derived experimentally by Kelly [12]. Instead of experimenting with flickering spatial patterns, Kelly measured contrast sensitivity at several fixed velocities for traveling waves of various spatial frequencies. Kelly used the well-known relationship of equivalence between the visual patterns flickering with temporal frequency  $\tau$ , and the corresponding steady patterns moving along the image plane with velocity  $\vec{v}$  such that [34]:

$$\tau = v_x \varrho_x + v_y \varrho_y = \vec{v} \cdot \vec{\varrho} \quad (1)$$

where  $v_x$  and  $v_y$  denote the horizontal and vertical components of the velocity vector  $\vec{v}$ , which is defined in the image plane  $xy$ , and  $\varrho_x$  and  $\varrho_y$  are the corresponding components of the spatial frequency  $\vec{\varrho}$ . Kelly found that the constant velocity CSF curves have a very regular shape at any velocity greater than about 0.1 degree/second. This made it easy to fit an analytical approximation to the contrast sensitivity data derived by Kelly in the psychophysical experiment. As a result, Kelly obtained the spatiovelocity

CSF, which he was able to convert into the spatiotemporal CSF using equation (1).

We use the spatiovelocity CSF model provided by Daly [4], who extended Kelly’s model to accommodate for the requirements of current CRT display devices (characterized by the maximum luminance levels of about  $100\text{ cd/m}^2$ ), and obtained the following formula:

$$CSF(\varrho, v) = c_0(6.1 + 7.3|\log(\frac{c_2 v}{3})|^3)c_2 v(2\pi c_1 \varrho)^2 \exp(-\frac{4\pi c_1 \varrho(c_2 v + 2)}{45.9}) \quad (2)$$

where  $\varrho = \|\vec{\varrho}\|$  is the spatial frequency in cycles per degree,  $v = \|\vec{v}\|$  is the retinal velocity in degrees per second, and  $c_0 = 1.14$ ,  $c_1 = 0.67$ ,  $c_2 = 1.7$  are coefficients introduced by Daly. In [4, 21] a more extended discussion on estimates of the retinal velocity is available, which takes into account the eye natural drift, smooth pursuit, and saccadic movements. Figure 1 depicts the spatiovelocity CSF model specified in equation (2).

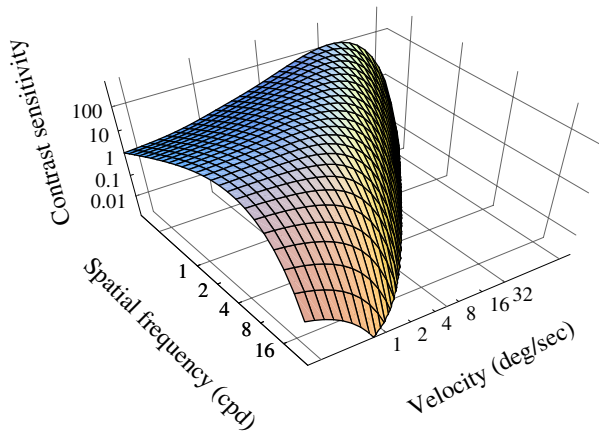


Figure 1: Spatiovelocity Contrast Sensitivity Function.

Although the spatiotemporal CSF is used by widely known video quality metrics, we chose to include the spatiovelocity CSF into our animation quality metric. Our design decision was encouraged by the observation that it is not clear whether the vision channels are better described as spatiotemporal (e.g., Hess and Snowden [10], and many other results in psychophysics) or spatiovelocity (e.g., Movshon et al. [20], and many other results especially in physiology). Also, accounting for the eye movements is more straightforward for a spatiovelocity CSF than for a spatiotemporal CSF [4]. Finally, the widely used spatiotemporal CSF was in fact derived from Kelly’s spatiovelocity CSF, which was measured for moving stimuli (traveling waves). However, the main reason behind our choice of the spatiovelocity CSF is that in our application we deal with synthetic animation sequences for which it is relatively easy to derive the velocity of moving spatial patterns across the image plane. In the following section we discuss this problem in detail.

## 3.2 Estimating Velocity of Moving Image Patterns

The 3D warping technique [18] which is traditionally used in the IBR applications has another interesting application. The 3D warping relies on the reprojection of every pixel in the reference image to its new location in the desired image (Figure 2). The 3D warping requires the depth (range) data that are registered with every image to properly solve occlusions which arise when the camera translational motion is involved. Acquiring such data with high precision is trivial for synthetic animation sequences, while it might be quite difficult task for video sequences representing real-world scenes. As a result of the 3D warping of a selected frame to the previous (following) frame in the camera animation sequence, the displacement vector between positions of the corresponding pixels which represent the same scene detail can be derived (refer to Figure 2). Because the time span between the subsequent animation frames is known (e.g., in the PAL composite video standard 25 frames per second are displayed), it is easy to compute the velocity vector based on the corresponding displacement vector. A vector field of pixel velocities defined for every image in the animation sequence is called the Pixel Flow (PF) which is the well-known notion in the digital video and computer vision communities [28]. Thus, for walk-through animation sequences that deal exclusively with changes of camera parameters a PF of good accuracy can be derived using the computationally efficient 3D warping technique<sup>1</sup>.

In the more general case of scene animation involving moving objects, the PF can be computed based on the scripts describing motion of characters, changes of their shape, and so on [27]. In this research, for the sake of efficiency reasons, we assume that the motion of animated objects is fully compensated by the smooth pursuit eye motion, 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 [23, 11], which means that efficiency of the eye tracking for such objects is very high. Thus, the computation of the PF for moving objects is not required, and the significant cost of such a computation [38] can be avoided.

## 3.3 Animation Quality Metric Algorithm

As the framework of our animation quality metric (AQM) we decided to expand the VDP [3]. The predictor was extensively tested through psychophysical experiments [5], and its integrity was shown for various contrast and visual masking models [7]. Furthermore, we found that the re-

<sup>1</sup>For the natural image sequences (video) sufficient spatial image gradients must exist to detect pixel displacements, in which case so called optical flow can be computed [28]. The optical flow computation is usually far less accurate and more costly than the PF computation for synthetic sequences.

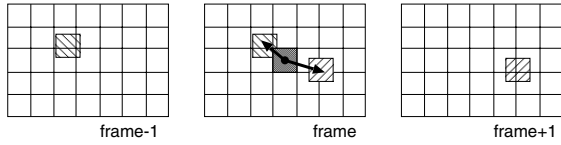


Figure 2: Displacement vectors for a pixel of the current frame in respect to the previous (frame-1) and following (frame+1) frames in an animation sequence. All marked pixels depict the same scene detail.

sponses of this predictor in graphics applications are very robust [17], and its architecture was suitable for an incorporation of the spatiovelocity CSF.

Figure 3 illustrates the processing flow of the AQM. Two comparison animation sequences are provided as input. For every pair of input frames the probability map  $M_{AQM}$  of perceiving the differences between these frames is generated as output.  $M_{AQM}$  provides for all pixels the probability values, which are calibrated in such a way that 1 Just Noticeable Differences (JND) unit [15, 3] corresponds to a 75% probability that an observer can perceive the difference between the corresponding image regions. While  $M_{AQM}$  provides local information on the differences, for some applications it is more convenient to use just a single value which measures the differences globally. We assumed that the percentage of pixels in  $M_{AQM}$  with the predicted differences over the 1 JND unit is a good measure of such global differences.

In the AQM computation each input frame undergoes the identical initial processing. At first, the original pixel intensities are compressed by the amplitude non-linearity and normalized to the luminance levels of the CRT display. Then the resulting images are converted into the frequency domain, and decomposition into spatial and orientation channels is performed using the Cortex transform which was developed by Daly [3] for the VDP. Then, the individual channels are transformed back to the spatial domain, and contrast in every channel is computed (the global contrast definition [7] with respect to the mean luminance value of the whole image was assumed).

In the next stage, the spatiovelocity CSF is computed according to the model of Kelly. The contrast sensitivity values are calculated using equation (2) for the center frequency  $\rho$  of each frequency band of the Cortex transform. The visual pattern velocity is estimated based on the average PF magnitude between the currently considered frame, and the previous and following frames (refer to Figure 2). As we discussed in Section 3.2, the PF can be estimated rapidly using the 3D warping technique, which requires access to the range data of the current frame and the camera parameters for all three involved frames. This means that the access to well localized data in the animation sequence is required. Since the visual pattern is maximally blurred in the direction of retinal motion, and spatial acuity

is retained in the direction orthogonal to the retinal motion direction [6], we project the retinal velocity vector onto the direction of the filter band orientation. The contrast sensitivity values are used to normalize the contrasts in every spatial frequency-orientation channel into the JND units. Next the visual masking is modeled using the threshold elevation approach [7]. The final stage is error pooling across all channels.

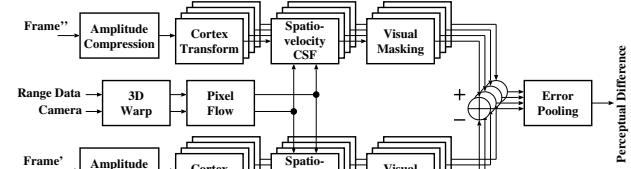


Figure 3: Animation Quality Metric. The spatiovelocity CSF requires the velocity value for every pixel, which is acquired from the PF. The PF is computed for the previous and following frames along the animation path in respect to the input frame' (or frame'' which should closely correspond to frame').

The AQM is well suited to computer graphics applications, and can be used to determine when a lower image quality will be not perceived for a given frame, and its local regions. In the following section we discuss the AQM application to guide rendering in high-quality walkthrough animation.

## 4 Static Environments - Walkthrough Animation

For animation techniques relying on keyframing the rendering costs depend heavily upon the efficiency of inbetween frame computation because the inbetween frames usually significantly outnumber the keyframes. We use IBR techniques [18, 16] to derive the inbetween frames. Our goal is to maximize the number of pixels computed using the IBR approach without deteriorating the animation quality as perceived by the human observer.

The quality of pixels derived using IBR techniques is usually lower than ray-traced pixels, e.g., in the regions of derived frames which are expanded in respect to the reference frames. The HVS sensitivity is especially high for such quality degradations when the PF values are low. We replace IBR-derived pixels in such regions with ray-traced pixels. The replacement is performed when the PF velocity is below a specified threshold value, which we estimated in subjective and objective (using the AQM) experiments [21]. In typical animations, usually only a few percent of the pixels are replaced, unless the camera motion is very slow. Those pixels are usually grouped around a so called focus of expansion [28] which represents the position in the image corresponding to the point towards which the camera is moving.

Since specular effects are usually of high contrast and they attract the viewers attention when looking at a video sequence [24], special care is taken to process them properly. Existing IBR methods require costly preprocessing to obtain specular effects of good quality [13, 9, 19, 14]. For example, a huge number of pre-calculated images is needed to obtain crisp mirror reflections. Because of these problems we decided to use ray tracing for pixels depicting objects with strong specular properties. We use our AQM to decide for which objects with glossy reflectance or transparent properties such computations are required.

Pixels representing objects in the inbetween frames which are not visible in the keyframes cannot be properly derived using the IBR techniques, and we apply ray tracing to fill the resulting holes in frames. An appropriate selection of keyframes is an important factor in reducing the number of pixels which must be ray traced.

In this work for the sake of simplicity we assume that initially the keyframes are placed sparsely and uniformly along the animation path which is known in advance. Then adaptive keyframe selection is performed, which is guided by the AQM predictions. We provide a detailed description of this solution in the following section. Then, we discuss the performance of our approach for a case study walkthrough animation.

## 4.1 Adaptive Refinement of Keyframe Placement

At first the initial keyframe placement is decided by choosing the constant length of  $N + 1$  frames for all animation segments  $S$ . Then every  $S$  is processed separately applying the following recursive procedure:

1. Generate the first frame  $k_0$  and the last frame  $k_N$  in  $S$  using ray tracing. The keyframes that are shared by two neighboring segments are computed only once.
2. Apply 3D warping to keyframes  $k_0$  and  $k_N$  to derive two instances  $k'_{[N/2]}$  and  $k''_{[N/2]}$  of an inbetween frame  $[N/2]$ .
3. Use the AQM to compute the probability map  $P_{Map}$  with perceivable differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$ .
4. Mask out from  $P_{Map}$  all pixels that must be ray traced because of the IBR deficiencies (discussed in Section 4). The following order for masking out pixels is taken:
  - (a) Mask out from  $P_{Map}$  pixels with low PF values (in [21] we discuss experimental derivation of the PF threshold value used for such masking).
  - (b) Mask out from  $P_{Map}$  pixels depicting objects with strong specular properties (i.e., mirrors, transparent and glossy objects). The item buffer [36] of frame  $k_{[N/2]}$  is used to identify pixels

representing objects with such properties. Only those specular objects are masked out for which the differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$  as reported in  $P_{Map}$  can be readily perceived by the human observer. In Section 4.2 we provide details on setting the thresholds of the AQM response, which are used by us to discriminate between the perceivable and imperceivable differences.

- (c) Mask out from  $P_{Map}$  holes composed of pixels that could not be derived from keyframes  $k_0$  and  $k_N$  using 3D warping.
5. If masked-out  $P_{Map}$  shows the differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$  for a bigger percentage of pixels than the assumed threshold value:
- (a) Split  $S$  at frame  $k_{[N/2]}$  into two subsegments  $S_1$  ( $k_0, \dots, k_{[N/2]}$ ) and  $S_2$  ( $k_{[N/2]}, \dots, k_N$ ).
  - (b) Process recursively  $S_1$  and  $S_2$ , starting this procedure from the beginning for each of them.
- Else
- (a) Composite  $k'_{[N/2]}$  and  $k''_{[N/2]}$  with correct processing of object occlusions [16, 26] to derive  $k_{[N/2]}$ .
  - (b) Ray trace all pixels which were masked out in the step 4 of this procedure, and composite these pixels with  $k_{[N/2]}$ .
  - (c) Repeat the two latter steps for all remaining inbetween frames, i.e.,  $k_1, \dots, k_{[N/2]-1}$  and  $k_{[N/2]+1}, \dots, k_{N-1}$  in  $S$ .

To avoid image quality degradation resulting from multiple resamplings, the fully ray-traced reference frames  $k_0$  and  $k_N$  are always warped in step 5c to derive all inbetween frames in  $S$ . Pixels to be ray traced, i.e., pixels with low PF values, pixels depicting specular objects with visible differences (such objects are selected once for the whole  $S$  in step 4b), and pixels with holes resulting from the IBR processing must be identified for every inbetween frame separately.

We evaluate the AQM response only for frame  $k_{[N/2]}$ . We assume that derivation of  $k_{[N/2]}$  applying the IBR techniques is the most error-prone in the whole segment  $S$  because its arclength distance along the animation path to either the  $k_0$  or  $k_N$  frames is the longest one. This assumption is a trade off between the time spent for rendering and for the control of its quality (we discuss the AQM costs in Section 5.2), but in practice, it holds well for typical animation paths.

Figure 4 summarizes the computation and compositing of an inbetween frame. We used a dotted line to mark those processing stages that are performed only once for

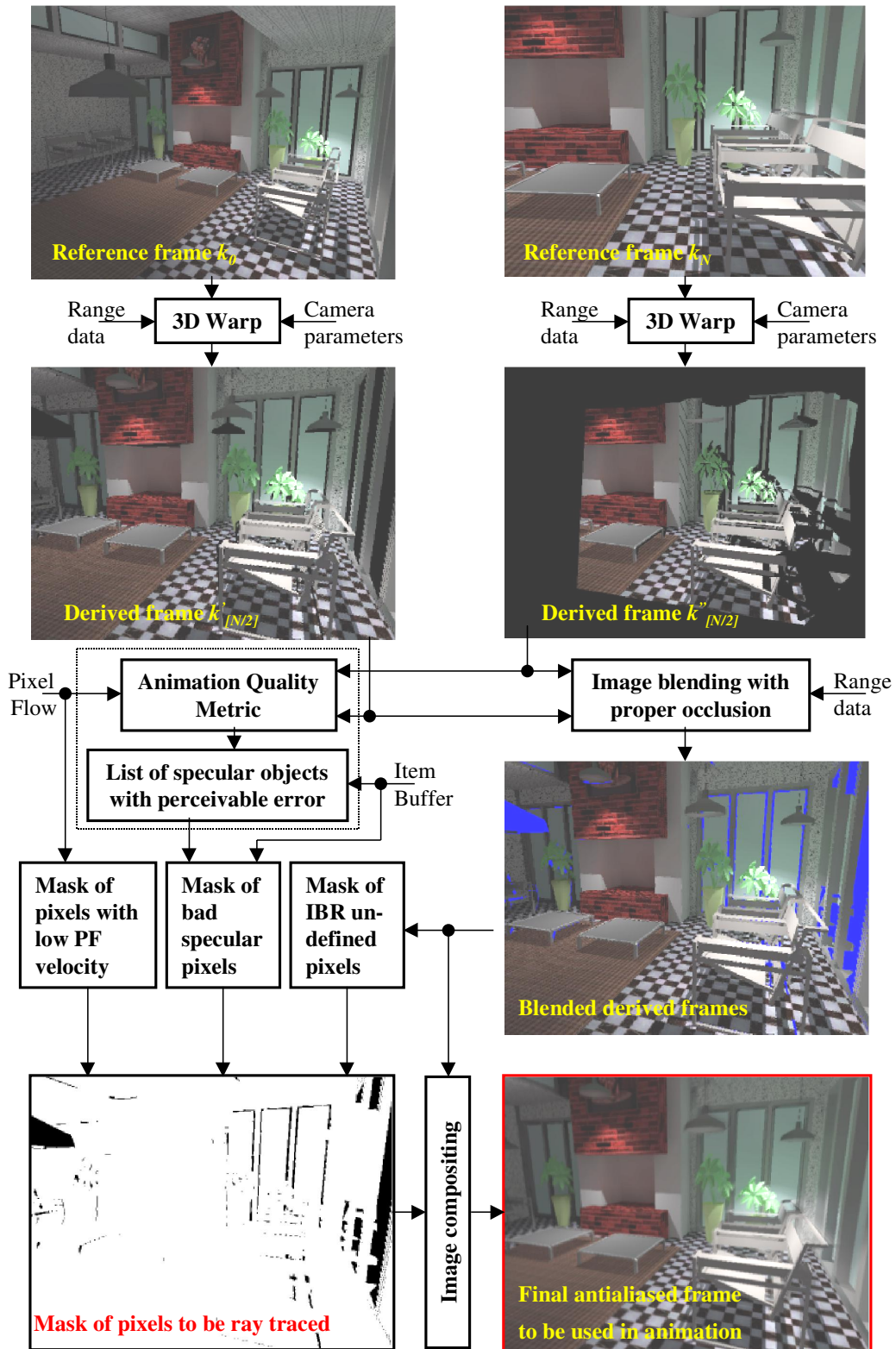


Figure 4: The processing flow for inbetween frames computation.



segment  $S$ . All other processing stages are repeated for all inbetween frames.

As a final step, we apply a spatiotemporal anti-aliasing technique, which utilizes the PF to perform motion-compensated filtering (refer to [21] for more details). On the Web page located under the URL: <http://www.mpi-sb.mpg.de/resources/aqm>, we provide the walkthrough animation sequences which result from our techniques of adaptive keyframe selection guided by the AQM predictions.

## 4.2 A Case Study Walkthrough Animation

In this work we choose as a case study a walkthrough animation for the ATRIUM scene scene shown in Figure 5 (more details on this scene are provided on the Web <http://www.mpi-sb.mpg.de/resources/atrium>). The main motivation for this choice were the interesting occlusion relationships between objects which are challenging for IBR. Also, a vast majority of the surfaces exhibit some view-dependent reflection properties, including the mirror-like and transparent surfaces, which made the computation of inbetween frames more difficult. Under such conditions, the AQM guided selection of keyframes and glossy objects within inbetween frames to be recomputed was more critical, and wrong decisions concerning these issues could be easy to perceive.

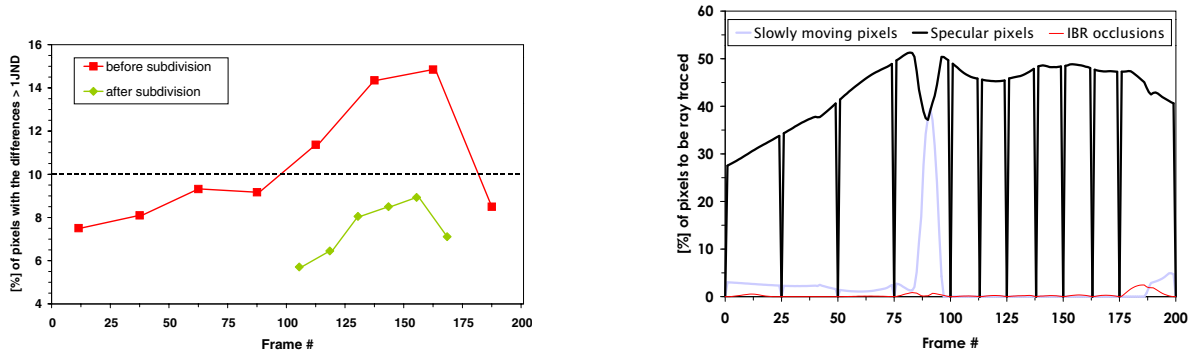


Figure 5: An atrium of the Research Quadrangle at the University of Aizu: rendering based on the simplified atrium model.

For our experiments we selected a walkthrough sequence of 200 frames. At the initial keyframe selection step, we assumed the length  $N + 1 = 25$  frames for each animation segment  $S$ . Figure 6a illustrates adaptive refinement of the initial keyframe placement guided by the AQM predictions. We use the global measure (refer to Section 3.3) of the differences between frames, i.e., the percentage of pixels in  $P_{Map}$  for which the differences are over 1 JND. Note that only pixels to be derived using the

IBR approach are considered, while pixels to be ray traced are masked out (refer to Section 4.1). The filled squares in Figure 6a show the global AQM predictions of the differences between the subsequent keyframe pairs:  $k_0$  warped to  $k'_{[N/2]}$ , and  $k_N$  warped to  $k''_{[N/2]}$  for every initial segment  $S$ . Segments with global predictions over 10% are split, and the filled diamonds show the corresponding reduction of the predicted perceivable differences between the newly inserted frames. The 10% threshold was chosen experimentally, and can be justified by the fact that for an animated sequence the observer can only fixate at one location per frame. For such a location and its surround of approximately 1 visual degree the eye sensitivity is high due to the foveal vision [32], while it decreases significantly for the remaining image regions which are perceived by means of the peripheral vision (eccentricity effect). The AQM sensitivity is tuned for the foveal vision because at the stage of animation rendering it is not known where the observer will be looking. This means that the AQM predictions might be too conservative for many image regions, and the degradation of quality usually cannot be perceived unless the observer attention is specifically attracted to these regions. To improve the rendering performance, we chose a trade-off solution in which visible differences between warped keyframes are allowed for a small number of pixels (up to 10%). Although some perceivable quality problems may arise for these pixels, it is most likely that the observer will not notice them at all.

The overall costs of the computation of inbetween frames are strongly affected by the number pixels of that must be ray traced. As we discussed in Section 4, we replace IBR-derived pixels by ray traced pixels for image patterns moving with low velocity. The graph in Figure 6b shows the percentage of such pixels for which the PF values are below the experimentally derived threshold velocity 0.5 degree/second (for details concerning the derivation of this threshold value refer to [21]). Also, we use ray tracing to derive pixels depicting specular objects for which the IBR technique leads to the AQM predicted degradation of the animation quality. As described in Section 4.1, for every segment  $S$  we run the AQM once to decide upon the specular objects which require recomputation. If a group of connected pixels representing an object (or a part of an object) exhibits differences greater than 2 JND units (a 93.75% probability of the difference discrimination), we select such an object for recalculation. If differences below 2 JND units are reported for an object by the AQM then we estimate the ratio of pixels exhibiting such differences to all pixels depicting this object. If the ratio is bigger than 25%, we select such an object for recomputation - 25% is an experimentally selected trade-off value, which makes a reduction in the number of specular objects requiring recomputation possible, at the expense of some potentially perceivable image artifacts. The graph in Figure 6b shows the percentage of specular pixels selected for recomputation. Finally, the percentage of pixels that are ray traced due to IBR occlusion problems is included



a)

b)

Figure 6: ATRIUUM walkthrough statistics: a) the AQM prediction of the perceived differences between the warped images of two neighboring reference frames, b) the percentage of pixels to be recalculated by ray tracing. In a) lines connecting the symbols were added for the figure readability and they do not have any meaning for unmarked frames.

in this graph. Table 1 summarizes the results depicted in Figure 6b by providing the average percentage of pixels per frame to be ray traced. Note that the keyframe pixels, which are always ray traced, are included in the average. The overall rendering time was reduced almost twofold. Even better performance can be expected for environments in which specular objects are depicted by a smaller percentage of pixels, and camera motion is faster.

In the following section we discuss another AQM application in which the metric is applied to guide the indirect lighting computation.

## 5 Dynamic Environments - Indirect Lighting Solution

An important research goal is to improve the performance of global illumination computation for animated sequences of high quality. This can be achieved by exploiting the temporal coherence in indirect lighting distribution in a better way. As a framework for global illumination computation, we chose the Density Estimation Photon Tracing (DEPT) algorithm [31]. 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 (refer to Figure 7). 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.

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 re-

duce 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 neighboring 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 [22], 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.

### 5.1 AQM Application

The main goal of adjusting the number of photons per frame  $N_{\text{frame}}$  is to keep the noise below a perceivable level. Increasing  $N_{\text{frame}}$  for each frame in a given animation segment is an expensive operation and should only be performed if the noise can be 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  $N_{\text{frame}}$  to minimize the computational costs without degrading the animation quality. The AQM requires two animation frames  $I_1$  and  $I_2$  as input, and



Table 1: Statistics of the ray traced pixels in the ATRIUM walkthrough.

Slow motion [%]	Specular objects [%]	IBR occlusions [%]	Keyframes [%]	Total [%]
2.4	40.8	0.3	6.0	49.5

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 [22] is used for the  $I_1(K)$  and  $I_2(K)$  computation as well.

As the result of AQM processing a map  $M_{\text{AQM}}$  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  $d_{\text{AQM}}$  of  $M_{\text{AQM}}$  pixels with differences over one unit Just Noticeable Difference (JND) is assumed [21]. The user chooses a certain threshold value  $d_{\text{thr}}$  of the AQM predicted differences, and when  $d_{\text{AQM}} > d_{\text{thr}}$ ,  $N_{\text{frame}}$  is doubled and the whole procedure is repeated until  $d_{\text{AQM}} < d_{\text{thr}}$ .

Since the same mesh is used for lighting reconstruction in frames  $I_1(K)$  and  $I_2(K)$ , the AQM is used to measure the perceivable differences between two equally biased indirect lighting solutions. Thus, the discretization error is the same for frames  $I_1(K)$  and  $I_2(K)$ , which means that all measured differences between the frames result from the stochastic noise. 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 quality of the indirect lighting reconstruction due to limitations in the spatial mesh resolution.

## 5.2 Issues of AQM Processing Efficiency

To reduce the costs of HVS modeling the AQM processing is performed only once per segment for the central frame  $K$ . Thus, the  $N_{\text{frame}}$  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 was adjusted to reduce such differences using the energy-based error metrics.

Our indirect lighting solution requires only seconds or

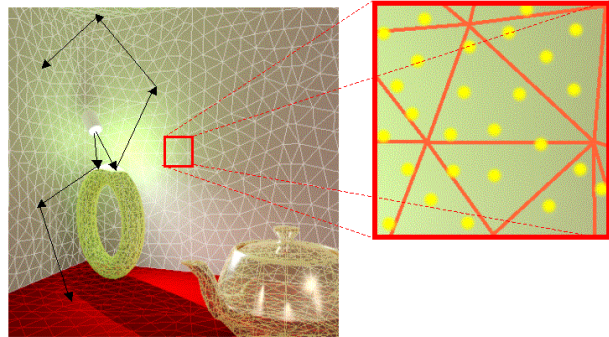


Figure 7: Density Estimation Photon Tracing algorithm: the lighting function is known implicitly as the density of photon hitting points over the scene surfaces.

single minutes to process complex scenes so the overhead of the HVS modeling plays an important role in our application of the AQM. 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 [23]. Also, we scale down images which are input to the AQM to  $256 \times 256$  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. All those simplifications result in the AQM processing time of about 4 seconds for a pair of compared frames.

## 5.3 Case Study Animations

We present results that we obtained for the ROOM scene (about 5,300 mesh elements, refer to Figure 8a), and the resulting animations are available under the URL <http://www.mpi-sb.mpg.de/resources/dynenv>. Also, we briefly summarize the results obtained for another scene SALON (about 48,600 mesh elements), which are qualitatively very similar and therefore do not need to be discussed in full length.

Figure 8b summarizes the results obtained using the perception-based procedure of noise level estimation as described in Section 5.1. 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

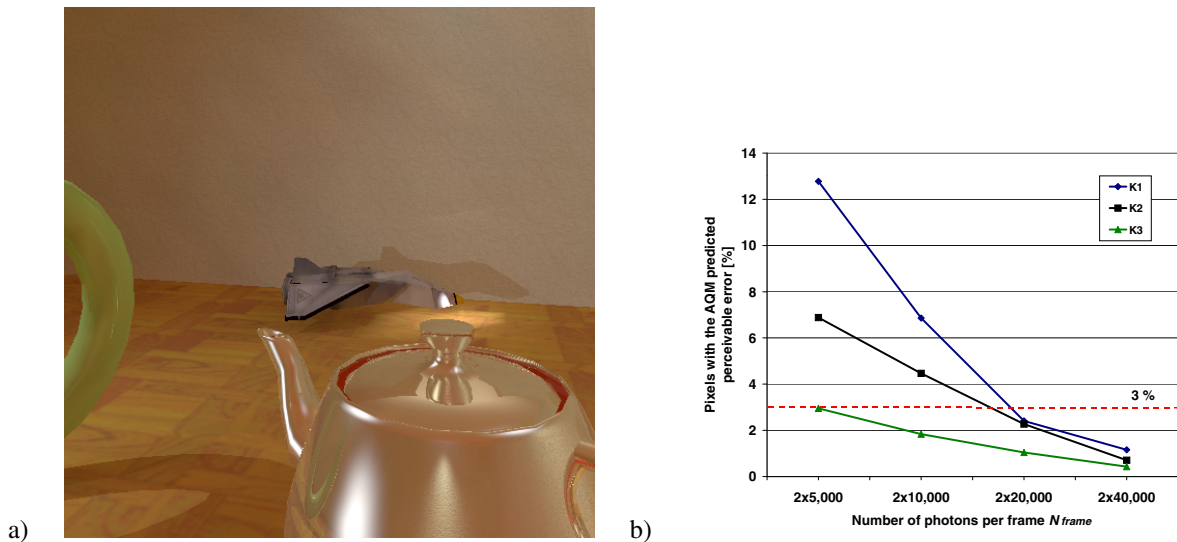


Figure 8: a) Example frame from the ROOM sequence. b) The AQM predicted percentage of pixels  $d_{\text{AQM}}$  with perceivable differences as a function of  $N_{\text{frame}}$  for the central segment frames  $K_i$ .

$N_{\text{frame}}$  chosen for every segment based on the graphs in Figure 8b. The average time of the indirect lighting computation per frame was 3.9 seconds.

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

Table 2: Final settings for the ROOM scene animation.

The results obtained for the SALON scene are very similar to the ones for ROOM. In general,  $N_{\text{frame}}$  fell into the range of 40,000–160,000 photons while the animation segment length lay between 15 and 44 frames. The average time of the indirect lighting computation per frame was 8.3 seconds. Figure 9a shows an animation frame obtained using our temporal photon processing algorithm. Figure 9b 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.

## 6 Conclusions

We showed two successful examples in which the perception-based guidance of rendering resulted in significant improvement of the computation performance. First, we proposed an efficient approach for rendering of high quality walkthrough animation sequences. Our contribution is in developing a fully automatic, perception-based guidance of inbetween frame computation, which minimizes the number of pixels computed using costly ray tracing, and seamlessly (in terms of the perception of animated sequences) replace them by pixels derived using inexpensive IBR techniques. Also, we discussed our global il-

lumination technique for dynamic environments which is suitable for high-quality animation rendering. We applied the AQM to decide upon the stopping conditions for our photon tracing procedures when the stochastic noise falls below the sensitivity level of the human observer. As a result the efficiency of computation is improved and the temporal aliasing is reduced with respect to traditional approaches which ignore temporal processing and perceptual considerations.

As future work we want to investigate our techniques in the context of MPEG coding. The accuracy of the rendering computation 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 image details from animations the compression performance can be improved.

## 7 Acknowledgments

We would like to thank Hans-Peter Seidel, Przemek Rokita, Takehiro Tawara, and Hiroyuki Akamine for their help in this research. Also, we thank Scott Daly for his stimulating comments on video quality metrics.

## References

- [1] A.A. Apodaca. Photosurrealism. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 315–322. Springer, 1998.
- [2] M.R. Bolin and G.W. Meyer. A perceptually based adaptive sampling algorithm. In *Proc. of Siggraph '98*, pages 299–310, 1998.

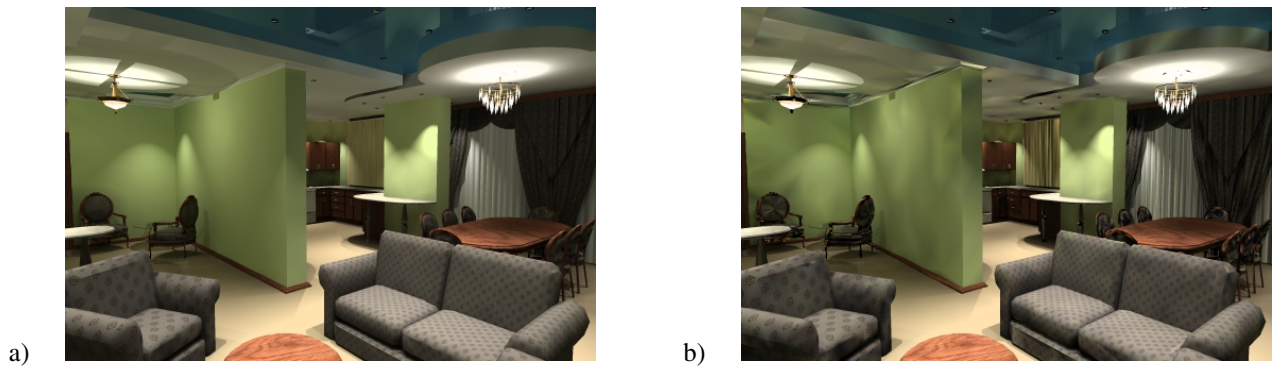


Figure 9: Example frame from the SALON sequence a) with temporal processing for  $N_{\text{frame}} = 80,000$  and the animation segment length 31 frames and b) without temporal processing for  $N_{\text{frame}} = 78,360$ .

- [3] S. Daly. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A.B. Watson, editor, *Digital Image and Human Vision*, pages 179–206. Cambridge, MA: MIT Press, 1993.
- [4] S. Daly. Engineering observations from spatio-velocity and spatiotemporal visual models. In *Human Vision and Electronic Imaging III*, pages 180–191. SPIE Vol. 3299, 1998.
- [5] S. Daly. Important issues for automating image quality estimation. *Siggraph '2000 Course Notes: Image quality metrics*, 44, July 2000.
- [6] M.P. Eckert and G. Buchsbaum. The significance of eye movements and image acceleration for coding television image sequences. In A.B. Watson, editor, *Digital Image and Human Vision*, pages 89–98. Cambridge, MA: MIT Press, 1993.
- [7] R. Eriksson, B. Andren, and K. Brunnstrom. Modelling of perception of digital images: a performance study. In *Human Vision and Electronic Imaging III*, pages 88–97. Proc. of SPIE Vol. 3299, 1998.
- [8] J.A. Ferwerda, S. Pattanaik, P. Shirley, and D.P. Greenberg. A model of visual masking for computer graphics. In *Proc. of Siggraph '97*, pages 143–152, 1997.
- [9] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen. The lumigraph. In *Proc. of Siggraph '96*, pages 43–54, 1996.
- [10] R.F. Hess and R.J. Snowden. Temporal properties of human visual filters: number shapes and spatial covariation. *Vision Research*, 32:47–59, 1992.
- [11] L. Itti, C. Koch, and E. Niebur. A model of Saliency-Based Visual Attention for Rapid Scene Analysis. *Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
- [12] D.H. Kelly. Motion and Vision 2. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10):1340–1349, 1979.
- [13] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of Siggraph '96*, pages 31–42, 1996.
- [14] D. Lischinski and A. Rappoport. Image-based rendering for non-diffuse synthetic scenes. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 301–314. Springer, 1998.
- [15] J. Lubin. A human vision model for objective picture quality measurements. In *Conference Publication No. 447*, pages 498–503. IEE International Broadcasting Convention, 1997.
- [16] W.R. Mark, L. McMillan, and G. Bishop. Post-rendering 3D warping. In *1997 Symposium on Interactive 3D Graphics*, pages 7–16. ACM Siggraph, 1997.
- [17] W.L. Martens and K. Myszkowski. Psychophysical validation of the Visible Differences Predictor for global illumination applications. In *IEEE Visualization '98 (Late Breaking Hot Topics)*, pages 49–52, 1998.
- [18] L. McMillan. *An Image-Based Approach to 3D Computer Graphics*. Ph.D. thesis, North Carolina University, Chapel Hill, 1997.
- [19] G. Miller, S. Rubin, and D. Poncelen. Lazy decompression of surface light fields for precomputed global illumination. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 281–292. Springer, 1998.
- [20] J.A. Movshon, E.H. Adelson, M.S. Gizzi, and W.T. Newsome. The analysis of moving visual patterns. In C. Chagas, R. Gattas, and C. Gross, editors, *Pattern recognition mechanisms*, pages 117–151. Rome: Vatican Press, 1985.

- [21] K. Myszkowski, P. Rokita, and T. Tawara. Perception-based fast rendering and antialiasing of walkthrough sequences. *IEEE Transactions on Visualization and Computer Graphics*, 6(4):360–379, 2000.
- [22] Karol Myszkowski, Takehiro Tawara, Hiroyuki Akamine, and Hans-Peter Seidel. Perception-guided global illumination solution for animation rendering. *Proceedings of SIGGRAPH 2001*, pages 221–230, August 2001.
- [23] W. Osberger. *Perceptual Vision Models for Picture Quality Assessment and Compression Applications*. Ph.D. Thesis, Queensland University of Technology, 1999.
- [24] W. Osberger, A.J. Maeder, and N. Bergmann. A perceptually based quantization technique for MPEG encoding. In *Human Vision and Electronic Imaging III*, pages 148–159. Proc. of SPIE Vol. 3299, 1998.
- [25] M. Ramasubramanian, S.N. Pattanaik, and D.P. Greenberg. A perceptually based physical error metric for realistic image synthesis. In *Proc. of Siggraph'99*, pages 73–82, 1999.
- [26] J.W. Shade, S.J. Gortler, L. He, and R Szeliski. Layered depth images. In *Proc. of Siggraph'98*, pages 231–242, 1998.
- [27] M. Shinya. Spatial anti-aliasing for animation sequences with spatio-temporal filtering. In *Proc. of Siggraph'93*, pages 289–296, 1993.
- [28] A.M. Tekalp. *Digital video Processing*. Prentice Hall, 1995.
- [29] X. Tong, D. Heeger, and C. van den Branden Lambrecht. Video quality evaluation using ST-CIELAB. In *Human Vision and Electronic Imaging IV*, pages 185–196. Proc. of SPIE Vol. 3644, 1999.
- [30] C.J. van den Branden Lambrecht. *Perceptual models and architectures for video coding applications*. Ph.D. thesis, Ecole Polytechnique Federal de Lausanne, 1996.
- [31] 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.
- [32] B.A. Wandell. *Foundations of vision*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1995.
- [33] G.J. Ward. The RADIANCE lighting simulation and rendering system. In *Proc. of Siggraph'94*, pages 459–472, 1994.
- [34] A.B. Watson. Temporal sensitivity. In *Handbook of Perception and Human Performance, Chapter 6*. John Wiley, New York, 1986.
- [35] A.B. Watson. Toward a perceptual video quality metric. In *Human Vision and Electronic Imaging III*, pages 139–147. Proc. of SPIE Vol. 3299, 1998.
- [36] H. Weghorst, G. Hooper, and D.P. Greenberg. Improved computational methods for ray tracing. *ACM Transactions on Graphics*, 3(1):52–69, January 1984.
- [37] S. Winkler. A perceptual distortion metric for digital color video. In *Human Vision and Electronic Imaging IV*, pages 175–184. Proc. of SPIE Vol. 3644, 1999.
- [38] Hector Yee, Sumanta Pattanaik, and Donald P. Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics*, 20(1):39–65, January 2001.
- [39] E.M. Yeh, A.C. Kokaram, and N.G. Kingsbury. A perceptual distortion measure for edge-like artifacts in image sequences. In *Human Vision and Electronic Imaging III*, pages 160–172. Proc. of SPIE Vol. 3299, 1998.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.  
Spring Conference on Computer Graphics, Budmerice, Slovakia  
Copyright © ACM 2002 1-58113-608-0/02/009...\$5.00