EUROGRAPHICS 2012 / P. Cignoni, T. Ertl
(Guest Editors)

Volume 31 (*2012*), *Number 2*

Crowd Light: Evaluating the Perceived Fidelity of Illuminated Dynamic Scenes

Adrian Jarabo[1], Tom Van Eyck[2], Veronica Sundstedt[3], Kavita Bala[4], Diego Gutierrez[1] and Carol O'Sullivan[2]

[1] Universidad de Zaragoza, [2] Trinity College Dublin, [3] Blekinge Institute of Technology, [4] Cornell University

Abstract

Rendering realistic illumination effects for complex animated scenes with many dynamic objects or characters is computationally expensive. Yet, it is not obvious how important such accurate lighting is for the overall perceived realism in these scenes. In this paper, we present a methodology to evaluate the perceived fidelity of illumination in scenes with dynamic aggregates, such as crowds, and explore several factors which may affect this perception. We focus in particular on evaluating how a popular spherical harmonics lighting method can be used to approximate realistic lighting of crowds. We conduct a series of psychophysical experiments to explore how a simple approach to approximating global illumination, using interpolation in the temporal domain, affects the perceived fidelity of dynamic scenes with high geometric, motion, and illumination complexity. We show that the complexity of the geometry and temporal properties of the crowd entities, the motion of the aggregate as a whole, the type of interpolation (i.e., of the direct and/or indirect illumination coefficients), and the presence or absence of colour all affect perceived fidelity. We show that high (i.e., above 75%) levels of perceived scene fidelity can be maintained while interpolating indirect illumination for intervals of up to 30 frames, resulting in a greater than three-fold rendering speed-up.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—;

1. Introduction

A large body of work in graphics is devoted to the problem of efficiently rendering realistic illumination in complex scenes with dynamics, but to what extent are approximations in global illumination perceptually noticeable? Most approximate rendering algorithms ignore perception, or use early vision based perceptual metrics to accelerate performance [BM98, MTAS01, DBD*07]. However, in the case where the illumination algorithm introduces approximations that are above threshold, there is not much understanding of how good (or bad) an algorithm is in terms of the image quality it achieves.

Recently there has been interest in understanding how approximations in global illumination are perceived. For instance, Yu and colleagues [YCK*09] characterize the effect of visibility approximations, while Křivánek et al. [KFB10] study the effect on image quality of approximations made by VPL-based rendering algorithms, like [Kel97], with clamping. But these approaches are based on static scenes, and the effect of dynamics on perceived accuracy is not considered.

In this paper we study a class of illumination approximations commonly used in rendering algorithms in production pipelines [PFHA10] for dynamic scenes. These approaches use spherical harmonics, and therefore typically produce low frequency illumination [RH01]. Several aspects of our problem formulation raise important questions, such as:

1. How noticeable are approximation errors in complex aggregates such as crowds?
2. Does the complexity of the entities in the crowd affect quality?
3. Does the type of crowd motion make the errors more or less perceivable?
4. Are errors in direct or indirect lighting more salient?
5. What effect does colour have on the perceived fidelity of illuminated crowd scenes?

We present a methodology to evaluate the perceived fidelity of illumination solutions in scenes with dynamic crowds. A series of psychophysical experiments are conducted to explore how a simple approach to approximating global illumination (GI), using interpolation in the temporal domain, affects the perceived fidelity of dynamic scenes with varying degrees of geometric, motion and illumination complexity. We are particularly interested in the problem of animating crowds of realistic human characters in complex scenes, and our results provide valuable insights into how errors introduced by different types of illumination approximations are perceived, and what factors affect them. For randomly moving crowds of human characters, for example, we were able to approximate illumination by simply interpolating indirect illumination for intervals *of up to 30 frames*, while maintaining high (i.e., above 75%) levels of perceived scene fidelity. But interpolating the direct lighting coefficients was much less effective. The answers to the above questions for animated human crowds are:

1. Errors in indirect illumination are more easily masked, allowing for coarser interpolation schemes.
2. Indirect illumination approximations are less noticeable for complex animated characters.
3. The more random and unstructured a motion is, the larger the approximation errors of indirect lighting can be.
4. Errors are more perceivable in direct illumination, specially for contact shadows.
5. The most acceptable approximation scheme for human crowds is to interpolate indirect illumination in colour scenes.

Finally, we compare our results with the outputs of a recent objective metric, and show that the predicted errors may indeed be masked by the dynamics of the aggregates, suggesting new avenues for exciting future work in this field. Our methodology has proven effective for evaluating the visual fidelity of complex dynamic scenes. It is our hope that these results will help to guide the development of faster rendering algorithms for complex dynamic scenes.

2. Related Work

Visual perception in computer graphics has received a lot of attention over the past few years [OHM*04, BCFW08, MMG11]. By understanding the limitations of the human visual system (HVS), rendering algorithms can be modified to eliminate unnecessary computations which will produce images with no perceivable difference to the observer. For instance, it is known that observers do not require a physically accurate simulation of the illumination in order to perceive a scene as realistic [MCTG00]. Common approaches use a perceptual metric to guide faster convergence in expensive global illumination algorithms (see for instance [MTAS01]), or exploit specific characteristics of the HVS, such as contrast and spatial masking [DBD*07]. Other approaches include visual attention to guide calculations, using less computations in visually unimportant regions, and leveraging aspects such as change blindness, saliency and task importance [FP03, SGA*07, HC09]. Yee et al. [YPG01] make use of the eye's reduced sensitivity to objects in motion in combination with a visual attention model to decide upon the quality of a given image pixel.

Other techniques are more specifically tailored to existing rendering algorithms and the specific illumination components they compute [SFWG04, DSSC05]. In this line, the work of Yu et al. [YCK*09] and Kozlowski and Kautz [KK07] show that perfect visibility is not necessary in indirect or glossy reflections, which can be used in interactive global illumination rendering [RGK*08].

Ramanarayanan et al. [RBF08] recently characterized the visual properties of general aggregates, and derived metrics to predict when static images of two different aggregates would have similar appearance. These metrics are used to substitute geometrically complex aggregates for simpler versions. This work is strongly related to ours: here we analyze the perceived fidelity of lighting in animations, for different types of *dynamic* aggregates. In the specific context of crowd rendering, there exist several works focusing on geometry simplification using different levels-of-detail (for a survey, see [RD05]). Additionally, several efforts focus on studying the perception of variety in a crowd, in order to create visually heterogeneous crowds from the combination of small sets of models and animations [MLD*08].

The seminal work on visual equivalency [RFWB07], shows that observers focus on the interaction of shape, material and illumination when deciding on the fidelity of a scene. The authors also evaluate and measure how illumination maps can be transformed without affecting the appearance of objects in images. Vangorp et al. [VCF*09] extend this work to dynamic scenes, analyzing what transformations on illumination maps preserve the appearance of objects in animations. Recently, the effect of visibility approximations has been characterized [YCK*09], while Křivánek et al. [KFB10] analyze different parameters in a global illumination approximation, and their influence when preserving the object's appearance.

Our work complements these recent studies, focusing on understanding the perceived fidelity of illumination in complex dynamic scenes, in particular crowds. Specifically, we consider the case of complex low-frequency illumination, with light coming from an environment map. This is described in the following section, along with our proposed lighting approximations, which will be used in the experiments to test the perceived fidelity of the scenes.

3. Illumination

Let us assume a scene is illuminated by an environment map. Light from a direction ω reflected at point x in direction ω_o can be modeled as:

$$L_o(x,\omega_o) = \int_{\Omega^+} L_i(x,\omega)\rho(x,\omega,\omega_o)V(x,\omega)(\omega\cdot n)d\omega \quad (1)$$

where L_i and L_o represent incident and outgoing radiance respectively at x, ρ is the BRDF, V is the visibility function and n is the normal at x. While Equation 1 only models direct illumination from distant light sources, it can be extended to account for the full light transport in a scene, assuming that $L_i(x,\omega)$ is a distant light source (i.e. it remains constant for all points x in the scene). We can rewrite Equation 1 as:

$$L_o(x,\omega_o) = \int_{\Omega^+} L_i(\omega) T(x,\omega,\omega_o) d\omega \qquad (2)$$

where the term $T(x,\omega,\omega_o)$ models the radiance transfer between the incoming and outgoing directions at x. This accounts for global illumination effects such as subsurface scattering or diffuse inter-reflections. Note that we refer to light coming directly from the environment map as *direct illumination*; when computing inter-reflections, we use the term *global illlumination*.

An efficient, well established way to approximate the radiance transfer is based on a spherical harmonics (SH) decomposition of lighting. It exploits the fact that the terms in Equation 2 are directionally dependent and vary smoothly, so they can be stored into a SH basis. Calculating L_o then becomes a fast dot product between the SH coefficients of each term (for a more detailed explanation, we refer the reader to some of the original publications on the topic [RH01, SKS02]).

3.1. Interpolating Spherical Harmonics

The main drawback of the SH technique is the massive computation required to pre-calculate the basis coefficients for the transfer function $T(x,\omega,\omega_o)$, due to the required extensive sampling of the light transport. To speed up calculations, we propose a simple approximation technique, based on interpolation in the temporal domain: for a few selected keyframes, T is calculated accurately; then, for each frame i between two keyframes $[kf_0, kf_1]$, T_i is obtained by linearly interpolating between T_{kf_0} and T_{kf_1}, such that each coefficient $T_i(c)$ in T_i is $T_i(c) = lerp(T_{kf_0}(c), T_{kf_1}(c), i)$.

In our psychophysical experiments, we will then study how far apart the keyframes can be (how much we can rely on simple interpolation) before an observer starts noticing illumination and shading artifacts. We also explored other interpolation schemes, such as cubic interpolation and cubic Hermite splines. However, we found that those methods not only have a higher computational cost, but they are also not necessarily more accurate. While our interpolation method is extremely simple and works only for pre-fixed animations, it suffices for the purpose of investigating lighting fidelity. Additionally, it allows us to exert direct control on the quality of final images by simply setting varying interpolation intervals, defined as the number of frames N between key frames.

4. Description of Experiments

Our aim is to study the perceived fidelity of illuminated dynamic scenes, and to explore the various factors that affect this perception. We therefore analyze the degree of simplification and inaccuracy that can be introduced in lighting computations before they are noticed by a human observer. This information can then be exploited in rendering algorithms to accelerate rendering computations. We focus specifically on the problem of rendering a crowd of characters, which is a particularly complex form of animated aggregate that poses a major challenge in many application areas, such as movies and games.

There are several factors that we hypothesize will affect the perceived fidelity of lighting in animated crowd scenes:

1. Interpolation Type (TYP): we have several choices of what coefficients to interpolate at each frame, and for what parts of the scene. We hypothesize that this decision will have different performance and perceptual implications, depending on the other factors being investigated.
2. Colour (COL): we want to examine the effects of illumination in the presence and absence of colour bleeding, as the latter may introduce additional visual cues that could impact perceived fidelity, again most likely in interaction with the other factors.
3. Character Object (OBJ): an animated human character will be perceived differently than a simple, non-animated object. Again, it is not clear whether this will cause errors to be more or less perceivable, so it is also likely to vary depending on the other factors. (It is worth noting again that the aim of our study is to identify the optimal conditions for rendering crowds of animated human characters).
4. Crowd Movement (MOV): the type of overall motion of the crowd affects how accurate the lighting is perceived to be. It was not clear whether more complex motion will mask or emphasize inaccuracies, so we hypothesize that this effect will vary depending on the other factors.
5. Interpolation Interval (INT): perceived fidelity will decrease with increasing interval size, as the illumination will be inaccurate for longer intervals.

To test these five hypotheses, we ran a set of psychophysical experiments. First, in Experiment 1, we explored the effect of interpolating the full radiance transfer matrix T, which provides a significant speed-up. However, this speed-up came at the cost of drastically reduced visual fidelity, immediately dropping to below 40% acceptability even when interpolating only every second frame. We therefore tested a second set of conditions, in Experiment 2, where we explored two different methods of interpolation, i.e., interpolating the lighting of only the crowd objects in the scene (TYP1) and interpolating only the indirect coefficients for the entire scene (TYP2). One group of participants viewed the animated scenes in greyscale only (No Colour) and a second group viewed animations with colour added to the environment (Colour). Our main finding is that TYP2 interpolation, i.e., interpolating only the indirect lighting coefficients for the full scene, provides the highest fidelity for

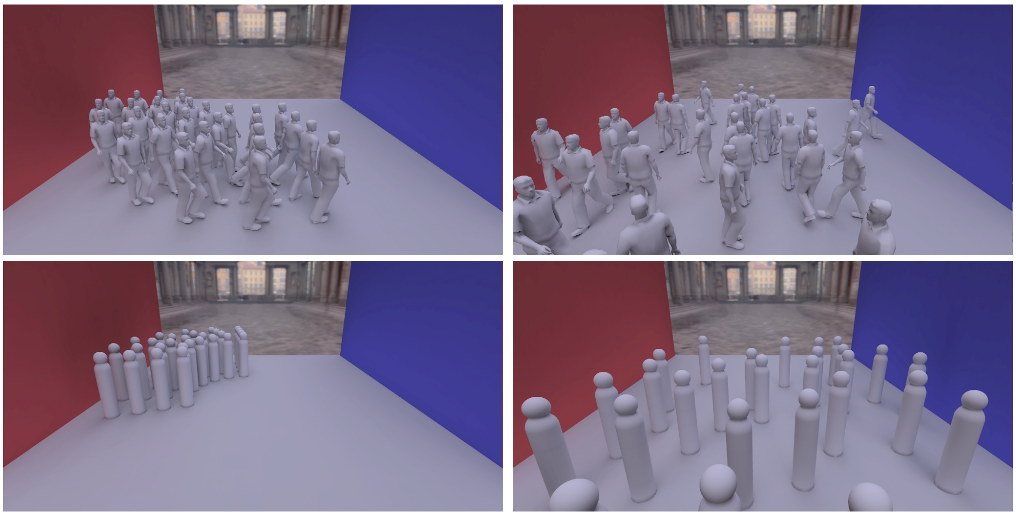


Figure 1: Example frames from the gold standard animations shown in our experiments, showing the two character object (OBJ) types (top: *Human*, bottom: *Pawn*), and the two movement (MOV) types (left: *Army*, right: *Random*).

scenes with animated humans, in particular when the environment includes colour and when the crowd is moving in a complex, random manner (see Section 5).

4.1. Method and Stimuli

Twenty participants volunteered for Experiment 1 (12M, 8F) and 40 participants for Experiment 2 (27M, 13F). Their ages ranged from 22-50, they had different educational backgrounds and all had normal acuity and color vision. The stimuli used in all experiments consist of pairs of short movies showing animated crowds. Each movie is eight seconds long, with a resolution of 1024 x 512 pixels, rendered with 4x antialiasing. Before the experiment starts, each participant is shown a gold standard animation for each different combination of factors, thereby familiarizing them with what a correct illumination solution will look like. For the second set of experiments, where the artifacts were more subtle, they were also shown examples of the types of errors to look out for.

We use a simple yes-no psychophysical task to test the perceived fidelity of each approximated scene to the gold standard. During each experiment, the participant is shown two videos running simultaneously on two different monitors. On one screen, the animation to be evaluated is displayed, where the matrix T is interpolated using the method described in Section 3.1, while on the other screen the *Gold Standard* (with T calculated accurately every frame) depicting the same factors is shown, but from a different point of view (to avoid direct side-by-side comparisons). The videos to be evaluated are played in random order, each with several repetitions (4 in Experiment 1, 2 in Experiment 2). The participants are informed which is the reference animation and which is the scene being evaluated. After each video pair terminates, a blank screen is shown for three seconds, during which the participant is required to answer 'Yes' or 'No' to the question: "is the illumination in the scene being evaluated the same quality as in the gold standard?". The gold standard is also compared with itself (i.e., interpolation interval size 1, but shown from the evaluation point of view) for 30% of the video pairs shown, in order to provide a baseline for comparison and to avoid a response bias (i.e., to ensure that there is a balance between 'Yes' and 'No' answers).

The scenario used for the animations is a ground plane bounded by two walls on the sides. These walls act as large occluders, limiting the incident direct lighting in the scene, while accentuating the effect of indirect illumination. The crowd depicted in each scene consists of 28 members, a number large enough to see the crowd as a whole, while allowing enough space to notice illumination interactions. The material chosen for both the crowd members and the scenario is purely Lambertian. Plain color is used as albedo, in order to avoid any possible masking effect introduced by textures, which could potentially hide illumination artifacts [FSPG97]. All videos are rendered using the spherical harmonics lighting method, which calculates the radiance transfer matrix T of the scene and uses it to integrate the illumination (see Section 3). The *uffizi* environment map from [Deb98] is used to illuminate the scenes. It has been chosen since it stores low-frequency illumination, and shows a plausible urban scenario for the crowds. Depending on the videos, the environment map can be rendered in one or three channels (gray-scale or RGB), but in both cases the light emitted is monochrome, i.e., luminances. This way, the incoming illumination in both the grey-scale and colour scenes

	Experiment 1	Experiment 2
OBJ	{*pawn, human*}	{*pawn, human*}
MOV	{*army, random*}	{*army, random*}
ILL	{*visibility only, full GI*}	*full GI*
TYP	full T	{TYP1, TYP2}
COL	*No Colour*	{*No Colour, Colour*}
INT	$\{1,2,3,4\}$	$\{1,2,5,10,30,60\}$

Table 1: Summary of the variables from the experiments: OBJ refers the objects used to form the crowd; MOV the types of movement; ILL the illumination setups; TYP the type of interpolation used; COL the color of the test animations; and INT the interpolation intervals used to approximate the illumination.

is the same, so the illumination colour in the scenes is never affected by the map.

4.2. Variables

In both experiments, we vary the type of objects used as crowd members, and the movement of the crowd (see Table 1 for a summary of the variables in the experiments). The first Object (OBJ) is an animated *Human* character, which is typical for the kinds of dynamic crowd scenes we are interested in accelerating. The second is a simplified, non-deformable *Pawn* character, included as a control for comparison purposes. While both objects have similar dimensions, they present two levels of complexity both in geometry and temporal behavior: the Pawn is a static smooth model while the animated Human provides sharp gradients and self-occlusions. We also explore two different types of Movement (MOV): *Army*, where the characters march in formation); and *Random*, where each member moves independently of the others, while always avoiding collisions. An aggregate whose members are uniformly distributed is perceived as less complex than another with a random distribution [Don06]. Similarly, if the crowd members move in a synchronized, structured pattern, they may be perceived as less complex than if moving with uncorrelated, random motion. Examples of both types of objects and movements can be seen in Figure 1).

In Experiment 1, we interpolate the full radiance transfer matrix T. Thus, the illumination is interpolated for all the geometry in the scene, and both the direct T_d and indirect T_{ind} components of T, such that $T = T_d + T_{ind}$. In this experiment, we also rendered the scenes using two different illumination configurations: the first one accounts for the direct *visibility only*, i.e. the $V(x,\omega)$ term in Equation 1, thus ignoring inter-reflections. The second one is a global illumination setup (*full GI*), with up to three light bounces taken into account. In both cases, the incoming light is approximated with 1024 samples over the hemisphere. Due to the amount of light lost when ignoring indirect illumination, the animations rendered with the *visibility only* configuration would in principle have less energy than the *full GI*. To compensate for this, the overall luminance in the *visibility only* scenes is normalized using the approach proposed by Křivánek et al. [KFB10], where the luminance is re-scaled such that the average luminance matches the average luminance in the *full GI* version of the scene. (Note that there was a different Gold standard for each of the two illumination types). Example frames from these scenes can be found in the supplementary material.

In Experiment 2, we explore how the perceived fidelity is affected by two different Interpolation Types (TYP). In TYP1, we interpolate the illumination for the objects in the crowd only i.e., both T_d and T_{ind} are interpolated only for the crowd's models, keeping accurate computations for the rest of the scene. This allows us to investigate the perception of interpolating lighting interactions between crowd members (e.g., shadows between models), and for each model with itself (e.g., self shadowing), together with the perceived consistency of shading in the full crowd. In TYP2, we interpolate only the indirect component T_{ind} of the radiance matrix T for the full scene (crowd and scenario) and keep T_d calculated every frame, as indirect illumination is known to be more easily approximated than direct illumination, given its lower-frequency nature.

For Experiment 1, we use a clear grey colour ($k_d = 0.67$) for all surfaces. However, colour bleeding in global illumination solutions can yield visual cues about orientation and three-dimensional layout [MTK*01]. Therefore, in Experiment 2, we explore the influence of colour (COL) as a between-groups factor, with the first group of 20 participants viewing the scene in grey-scale, or *No Colour*, and the second group viewing the scene with *Colour*, where the grey walls are replaced by one red ($k_d = 0.67, 0.28, 0.28$)) and one blue ($k_d = 0.28, 0.28, 0.67$) wall. The rest of the scene (the floor and the crowd) remains unchanged.

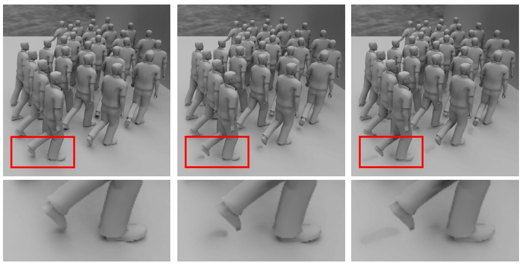


Figure 2: Comparison of a frame from three videos used in the first experiment (interpolating the full radiance transfer matrix), rendered using different interpolation intervals N: from left to right, $N = \{1,2,4\}$.

Finally, in Experiment 1 we use four values for the interpolation Interval (INT) N for the spherical harmonics coefficients. These intervals are $N = \{1,2,3,4\}$ frames, where $N = 1$ is the gold standard, i.e., no interpolation. Figure 2 shows a comparison between the results of different val-

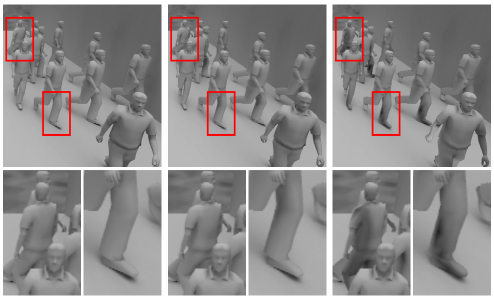


Figure 3: Comparison of a frame from three videos showing the same scene rendered using interpolation TYP1 with interpolation intervals $N = \{1,5,30\}$, from left to right. Each inset shows differences caused by interpolation in both self-shadowing and shadows cast by other crowd members.

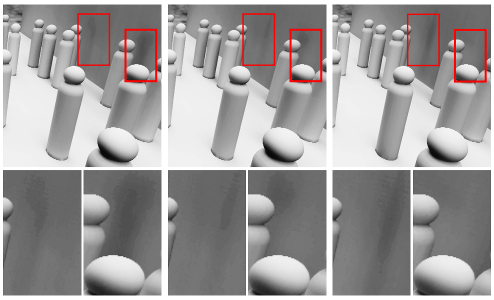


Figure 4: Comparison of a frame from three videos showing the same scene rendered using interpolation TYP2 with interpolation intervals $N = \{1,5,30\}$ (from left to right). Each inset shows differences produced by the interpolation and that appear in the indirect soft shadows. Note that the contrast and brightness has been increased for visualization purposes.

ues of N in one frame from the first experiment. For Experiment 2, as the artifacts are more subtle, we increase the number and range of interpolation intervals to INT $N = \{1,2,5,10,30,60\}$ (determined with pilot tests). Figure 3 shows an example of the effects that different N produce on the shading of crowd members using interpolation TYP1, while Figure 4 illustrates the artifacts that different N produce on indirect soft shadows using interpolation TYP2.

To summarise: in Experiment 1 we are testing 32 different combinations of within-subject variables (i.e., all participants saw all combinations): two types of OBJ (*Pawn*, *Human*), two MOV (*Army*, *Random*), two types of illumination (ILL) (*visibility only*, *full GI*), and four intervals INT. In each block of Experiment 2, we tested 48 combinations of within-subject variables: two OBJ, two MOV, two types of interpolation (TYP) and six INT, with COL (*No Colour*, *Colour*) as a between-groups factor, or categorical predictor (i.e., only half the participants saw one value or the other).

5. Experimental Results

In order to test our hypotheses outlined in Section 4, we must test for statistically significant differences in responses to the different stimuli. We are interested in both *Main Effects* (i.e., when a particular variable or factor has an overall effect, independently of the other variables); and *Interaction Effects* (i.e., when the effect of a variable differs depending on the level(s) of one or more of the other variables). To test for such effects, we use Repeated Measures Analysis of Variance (ANOVA) on the data from our psychophysical experiments. When we find main or interaction effects, we explore what is causing these effects further using a Neuman-Keuls post-hoc test for pair-wise comparisons of means. We only report effects that are significant at the 95% level, i.e., where the probability that the difference between means occurred by chance is less than 5% (i.e., $p < 0.05$).

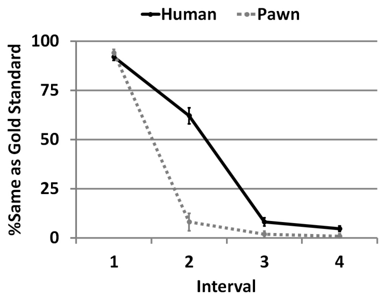


Figure 5: Results from Experiment 1, showing the interaction effect between Object (OBJ) and Interval (INT).

In Experiment 1, the variables to be tested are Illumination type *(ILL = visibility only, full GI)*, Object *(OBJ = Human, Pawn)*, *Movement (MOV = Army, Random)* and Interpolation Interval (INT = 1,2,3,4). We ran a four-way, repeated measures ANOVA with the four variables, and found main effects of Object OBJ ($F(1,19) = 43.2, p \approx 0$) and Interval INT ($F(3,57) = 439.53, p \approx 0$), and an interaction effect of Object and Movement OBJ*MOV ($F(3,57) = 59.504, p \approx 0$). Post-hoc tests revealed that the scenes with the *Human* object were perceived to be similar to the gold standard more often than those with the *Pawn* (42% vs. 26% overall), and that INT sizes 1, 2 and 3 were all significantly different from each other, except for the two largest, 3 and 4, which were both acceptable less than 5% of the time. However, post-hoc tests on the interaction effect reveal that only the Human object was in any way acceptable for any of the simplified scenes, and then only 60% of the time for the smallest interval size of 2 (see Figure 5). There was no effect of movement type. Nor did the illumination type have any effect, which indicates that attempting to interpolate the direct lighting coefficients (which was common to both types of illumination) gave rise to unacceptable artifacts in almost all cases. This was masked only 60% of the time by the complexity of the Human object, and only for the smallest interpolation interval (i.e., every second frame).

Therefore, in Experiment 2, we first tried interpolating the full lighting (i.e., both direct and indirect) for the objects in the crowd only, as the Human object's complexity appeared to have a masking effect in Experiment 1. Perhaps by restricting the interpolation to the Human model only, we could improve upon this result. We also tried interpolating the indirect lighting components of the entire scene only, as these are known to be easier to approximate. Therefore, the variables to be tested in this experiment, in addition to the same OBJ and MOV variables from Experiment 1, were Interpolation type *TYP1 = crowd objects only, TYP2 = indirect only*, and between-groups factor Colour type *(COL = No Colour, Colour)*. We therefore ran a five-way repeated measures ANOVA, with dependent (within group) variables TYP, OBJ, MOV and INT, and categorical predictor COL (between groups), followed by Neumann-Keuls post-hoc significance tests. All significant results are reported in Table 2, where the most interesting findings with respect to our motivation (i.e., rendering crowds of animated humans) are emphasized in bold.

We can see that there is only one main effect of interpolation level INT, as predicted, and the effects of all other variables are highly dependent on combinations of the others. There are several two-, three- and four-way interactions, and one five-way interaction, which vindicates our hypotheses that all of these factors affect the perceived fidelity of the lighting in animated crowd scenes. See Figure 6 for an overview of all our results. The most interesting findings for evaluating the perceived fidelity of illuminated scenes with animated humans is that approximating the indirect lighting in a scene is the best option for such scenarios: the humans look most realistic in this condition, in particular when they are undergoing random motion, and this effect is strongest in Colour scenes, i.e., exactly the kinds of scenes that are found in most applications that require realistically rendered crowds, such as movies and games. In summary, our five-way interaction effect shows that the perceived fidelity of dynamic colour scenes depicting animated human crowds is $>= 75\%$ up to an interpolation interval of 10 for structured army crowd movement, and 30 for the un-structured random motion. As we will see in the next section, this provides significant savings in computation for such scenes.

Comparison with a Video Quality Metric: We have also compared our findings with the recently published video quality metric of Aydin et al. [ACMS10] designed to predict visual differences in videos. Figure 7 shows a frame from the results of running the metric on two different scenes. The metric has been run on two videos for each scene, V_5 and V_{30}, with the illumination interpolated with $N = \{5, 30\}$ respectively. For the two scenes, the metric predicts clear visual differences between the reference V_{ref} and both interpolated videos. However, the results of our experiments show that illumination in V_5 is perceived as similar to V_{ref}. This shows that the perceived fidelity does not depend exclusively on the differences predicted by the metric. More work is needed in order to come up with a metric that predicts *perceived* quality for the case of complex dynamic aggregates.

Motion	Intp. Type	N	Time/frame	Speed-Up
Army	TYP1	5	4'18"	1.15x
Crowd	TYP1	5	4'18"	1.15x
Army	TYP2	10	1'36"	3.08x
Crowd	TYP2	30	1'21"	3.64x

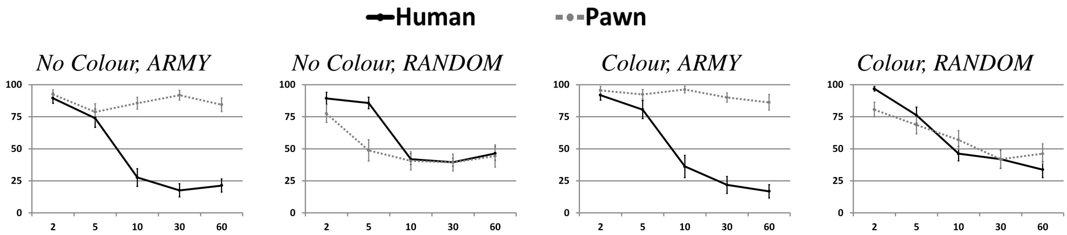
Table 3: Performance gain when precomputing human crowd scenes with color, for each motion and interpolation type, with the highest N accepted as as good as the reference more than 75% of the time.

Performance: Our results can be used to speed up rendering times while keeping approximation errors undetectable by observers. In our implementation, the SH coefficients are precomputed using a ray-tracer implemented in CUDA, with the animations being rendered in OpenGL. Since the cost of rendering is negligible compared to the precomputation, we focus here on the latter.

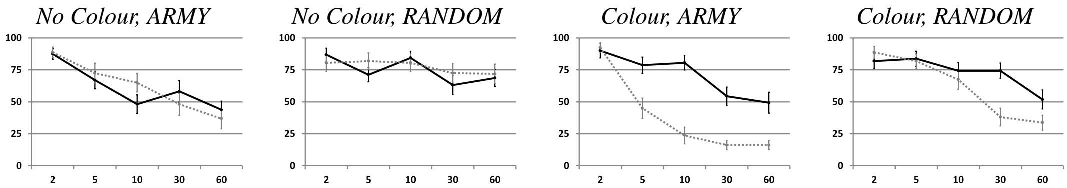
The SH coefficients are computed on a set of $1.75 \cdot 10^6$ total samples, of which 16% are placed on the crowd members. In the *full GI* lighting configuration, computing indirect illumination takes up to 75% of the total time. Table 3 shows the performance gained by using the largest interpolations that ensured that the scene is perceived as similar to the reference by more than 75% of the participants. The timings refer to the scenes with color, which we deem a more common rendering scenario. Measures are taken on an Intel Core I7 950@3.07GHz processor with an NVIDIA GeForce GTX470. Note that the speed-ups obtained for the cases shown (interpolating *only the members of the crowd* and *only the indirect illumination*) are bounded by 1.19x and 4x respectively, since they are only accelerating a portion of the total calculation (16% and 75% in each case).

Generalization: In our experiments only low-frequency illumination is considered. For high-frequency illumination, the approximation will fail, as seen in Experiment 1, where the artifacts in contact shadows were detected most times. However, if high-frequency illumination is introduced without approximation, low-frequency illumination might be approximated even more, since high-frequency lighting and shadows will probably be more salient, thus masking artifacts to some extent.

The experiments focus on a worst-case scenario, with a low appearance complexity of both the background and the elements of the crowd. Introducing more complexity in their appearance (more complex geometry, textures) would introduce visual masking, which might further hide artifacts produced by approximating the illumination. Additionally in this work neither saliency is taken into account, nor the psychological aspects of how observers tend to look at humans. Both might have a very important effect on the perception of visual fidelity, since probably they could make error sensitivity decrease in low-attended areas. Furthermore, our results



(a)Interpolation TYP1: coefficients of the crowd members only



(b)Interpolation TYP2: indirect illumination only for both crowd and scene

Figure 6: Results from Experiment 2: The x-axes show interpolation intervals, while the y-axes show the percentage of times the approximation was found to be of equal fidelity to the gold standard.

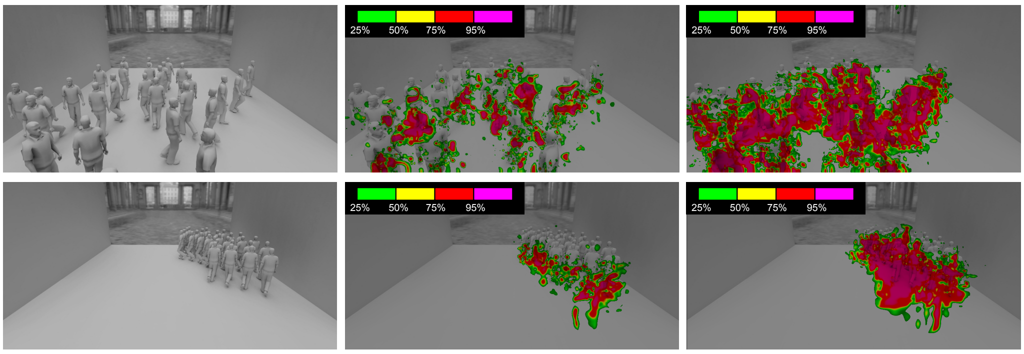


Figure 7: Visual differences predicted by the video quality metric of Aydin et al. [ACMS10] in two videos with different crowds between the reference video V_{ref} (left) and two approximations. The metric predicts visible differences when interpolating using interpolation TYP1 with both $N = 5$ (center) and $N = 30$ (right)), while in our experiments interpolating with $N = 5$ is perceived as similar to V_{ref}. Color scale means % of perceived difference.

suggest that the more complex the aggregate is, the more the illumination can be approximated. Intuitively, this would mean that increasing numerosity might allow increasing approximation levels without affecting perceived fidelity.

Finally, in our experiments the subjects are asked to perform a specific task (evaluating fidelity), and they are provided with a ground truth animation to compare with. In real settings (games, movies) the observers are probably not focused on finding errors, nor will they have a gold standard to guide them. So, the allowable approximation level would likely be higher in real applications.

6. Conclusions and Future Work

In this paper we have presented a methodology to evaluate how realistic lighting can be approximated using interpolation in complex dynamic scenes with crowd aggregates, without a decrease in perceived quality. A series of psychophysical experiments were conducted which explored the perceived fidelity of this approximation in scenes with varying degrees of geometric, motion and illumination complexity. Our results showed that errors in illumination could be masked by the aggregate characteristics. In our SH implementation, the illumination could, in some cases, be approximated at intervals of up to 30 frames without noticeable artifacts. In particular, our results show that coarser interpolation schemes could be used for complex aggregates where errors are masked more easily. Errors also become less noticeable as the complexity of the dynamic aggregates increases. The results also showed that the type of motion affects perception, so that it is possible to further approximate the illumination in the case of more random motion. Lastly, errors were easier to perceive in direct illumination and in

Effect	F-Test	Post-Hoc
Main Effects		
INT	$F(4,152) = 101.52, p \approx 0$	Perceived fidelity deteriorates with increasing INT, but 30 = 60
Two-way Interactions		
TYP*COL	$F(1,38) = 9.2, p < .005$	Interpolation Type (TYP) 2 beats TYP 1 for No Colour only
INT*COL	$F(4,152) = 3.4, p < .05$	No Colour better than Colour for INT=60 only
TYP*OBJ	$F(1,38) = 83.7, p \approx 0$	Pawn beats Human for TYP 1; **Human beats Pawn for TYP 2**
TYP*MOV	$F(1,38) = 92.0, p \approx 0$	Army beats Random for TYP 1; **Random beats Army for TYP 2**
OBJ*MOV	$F(1,38) = 33.5, p \approx 0$	Pawn beats Human for Army; **Human beats Pawn for Random**
TYP*INT	$F(4,152) = 4.8, p < .005$	TYP 1 deteriorates faster: INT 10 worse for TYP 1 than for TYP 2
OBJ*INT	$F(4,152) = 7.6, p < .00005$	Human beats Pawn at INT 5; Pawn beats Human for INT $>=$ 10
MOV*INT	$F(4,152) = 2.8, p < .05$	Army beats Random for INT 5; Random beats Army for INT 60
Three-way Interactions		
TYP*OBJ*COL	$F(1,38) = 23.6, p < .00005$	Pawn beats Human for TYP 1, Colour and No Colour; **Human beats Pawn for TYP2 and Colour**; equal for No Colour
OBJ*MOV*COL	$F(1,38) = 4.9, p < .05$	Pawn beats Human for Army only, Colour and No Colour
OBJ*INT*COL	$F(4,152) = 2.7, p < .05$	Human deteriorate faster for No Colour than for Colour; Pawn beats Human for INT $>=$ 10 and No Colour only
TYP*OBJ*MOV	$F(1,38) = 111.7, p \approx 0$	Pawn beats Human for TYP 1 and Army; **Human beats Pawn for TYP2 and Random**
TYP*OBJ*INT	$F(4,152) = 27.5, p \approx 0$	Human deteriorates faster than Pawn for TYP1, INT $>=$ 10; **Human beats Pawn for TYP2 and INT** $>=$ 10
TYP*MOV*INT	$F(4,152) = 6.8, p < .00005$	Army beats Random for TYP 1; but **Random beats Army for TYP 2 and INT** $>=$ 5
OBJ*MOV*INT	$F(4,152) = 4.0, p < .005$	Pawn beats Human for Army and INT $>=$ 10, not for Random
Four-way Interactions		
TYP*OBJ*INT*COL	$F(4,152) = 3.9, p < .005$	Pawn beats Human for TYP 1, both No Colour and Colour; **Human beats Pawn for TYP 2 with Colour**; equal for No Colour
OBJ*MOV*INT*COL	$F(4,152) = 3.39, p < .05$	Pawn beats Human for Army and No Colour (INT $>=$ 10); Pawn beats Human for Army and Colour (INT $>=$ 30); **Human beats or equals Pawn for Random in all cases**
TYP*OBJ*MOV*INT	$F(4,152) = 7.7, p < .00005$	Pawn beats Human for TYP 1 and Army (INT $>=$ 10) only; **Human beats or equals Pawn for TYP2 in all cases**
Five-way Interaction		
TYP*OBJ*MOV*INT*COL	F(4,152)=2.7, p<.05	Shown in Figure 6 and discussed in Section 5
Main finding: Perceived fidelity of Human is $>=$ 75% for TYP2 interpolation up to INT 30 for Random movement in Colour scenes		

Colour (between groups factor): COL 1 = No Colour (i.e., greyscale), COL 2 = Colour (i.e., coloured walls)
Interpolation type: TYP 1 = interpolate coefficients for crowd members only, TYP 2 = interpolate indirect coefficients only, but for full scene
Object types: OBJ 1 = animated Human character, OBJ 2 = Pawn control character
Movement types: MOV 1 = Army crowd motion, MOV 2 = Random crowd motion
Interval sizes: INT 2, 5, 10, 30, 60

Table 2: Significant results from the psychophysical experiments.

particular in contact shadows. Our work was also evaluated using an objective metric which showed that the predicted errors may indeed be masked by the dynamics of the aggregates.

This gained insight opens other avenues of future work. For instance, there are other properties of dynamic aggregates that can be explored, such as numerosity, variety, the effect of distance from the camera and level of detail or the appearance of the models within the aggregate. From a rendering perspective, it would be interesting to explore new approximation schemes, which would potentially yield further speed-ups. Last, our tests have shown a discrepancy between the prediction of an objective metric and the results of our experiments, which suggest that further work in this regard is still necessary.

Acknowledgments

The authors would like to thank Martin Prazak for providing the code for the random motion, the subjects for participation in the experiments, and the reviewers for their helpful comments. This work was partially sponsored by Science Foundation Ireland as part of the Metropolis project, by the European Commission, Seventh Framework Programme, through the projects GOLEM (Marie Curie IAPP, grant agreement no.: 251415) and VERVE (Information and Communication Technologies, grant agreement no.: 288914), by the Spanish Ministry of Science and Technology (TIN2010-21543), by the National Science Fundation under award CCF-0811680, and CAI through Programa Europa.

References

[ACMS10] AYDIN T. O., CADÍK M., MYSZKOWSKI K., SEIDEL H.-P.: Video quality assessment for computer graphics applications. *ACM Trans. Graph. 29* (2010), 161:1–161:12. 7, 8

[BCFW08] BARTZ D., CUNNINGHAM D., FISCHER J., WALLRAVEN C.: The Role of Perception for Computer Graphics . In *Eurographics 2008 - State of the Art Reports* (2008), pp. 59–80. 2

[BM98] BOLIN M. R., MEYER G. W.: A perceptually based adaptive sampling algorithm. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), pp. 299–309. 1

[DBD*07] DRETTAKIS G., BONNEEL N., DACHSBACHER C., LEFEBVRE S., SCHWARZ M., VIAUD-DELMON I.: An interactive perceptual rendering pipeline using contrast and spatial masking. In *Rendering Techniques (Proceedings of the Eurographics Symposium on Rendering)* (2007), Eurographics. 1, 2

[Deb98] DEBEVEC P.: Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of SIGGRAPH98, Computer Graphics Proceedings, Annual Conference Series* (1998), pp. 189–198. 4

[Don06] DONDERI D.: Visual complexity: a review. *Psychological Bulletin 132*, 1 (2006), 73–97. 5

[DSSC05] DEBATTISTA K., SUNDSTEDT V., SANTOS L. P., CHALMERS A.: Selective component-based rendering. In *Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia* (2005), GRAPHITE '05, pp. 13–22. 2

[FP03] FERWERDA J., PELLACINI F.: Functional difference predictors (fdps): measuring meaningful image differences. In *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on* (2003), vol. 2, pp. 1388 – 1392 Vol.2. 2

[FSPG97] FERWERDA J. A., SHIRLEY P., PATTANAIK S. N., GREENBERG D. P.: A model of visual masking for computer graphics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), SIGGRAPH '97, pp. 143–152. 4

[HC09] HASIC J., CHALMERS A.: Saliency in motion: selective rendering of dynamic virtual environments. In *Proceedings of the 2009 Spring Conference on Computer Graphics* (2009), SCCG '09, pp. 173–180. 2

[Kel97] KELLER A.: Instant radiosity. *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), 49–56. 1

[KFB10] KŘIVÁNEK J., FERWERDA J. A., BALA K.: Effects of global illumination approximations on material appearance. *ACM Trans. Graph. 29* (2010), 112:1–112:10. 1, 2, 5

[KK07] KOZLOWSKI O., KAUTZ J.: Is accurate occlusion of glossy reflections necessary? In *Proceedings of the 4th symposium on Applied perception in graphics and visualization* (2007), APGV '07, pp. 91–98. 2

[MCTG00] MCNAMARA A., CHALMERS A., TROSCIANKO T., GILCHRIST I.: Comparing real & synthetic scenes using human judgements of lightness. In *Proceedings of the Eurographics Workshop on Rendering Techniques 2000* (2000), pp. 207–218. 2

[MLD*08] MCDONNELL R., LARKIN M., DOBBYN S., COLLINS S., O'SULLIVAN C.: Clone attack! perception of crowd variety. *ACM Trans. Graph. 27* (2008), 26:1–26:8. 2

[MMG11] MCNAMARA A., MANIA K., GUTIERREZ D.: Perception in graphics, visualization, virtual environments and animation. In *ACM SIGGRAPH Asia 2011 Courses* (New York, NY, USA, 2011), SIGGRAPH Asia '11, ACM. 2

[MTAS01] MYSZKOWSKI K., TAWARA T., AKAMINE H., SEIDEL H.-P.: Perception-guided global illumination solution for animation rendering. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), pp. 221–230. 1, 2

[MTK*01] MADISON C., THOMPSON W., KERSTEN D., SHIRLEY P., SMITS B.: Use of interreflection and shadow for surface contact. *Perception & Psychophysics 63*, 2 (2001), 187–194. 5

[OHM*04] O'SULLIVAN C., HOWLETT S., MCDONNELL R., MORVAN Y., O'CONOR K.: Perceptually adaptive graphics. In *Eurographics state of the art reports* (2004), vol. 4, pp. 141–164. 2

[PFHA10] PANTALEONI J., FASCIONE L., HILL M., AILA T.: Pantaray: fast ray-traced occlusion caching of massive scenes. *ACM Trans. Graph. 29* (2010), 37:1–37:10. 1

[RBF08] RAMANARAYANAN G., BALA K., FERWERDA J. A.: Perception of complex aggregates. *ACM Trans. Graph. 27*, 3 (2008), 1–10. 2

[RD05] RYDER G., DAY A. M.: Survey of real-time rendering techniques for crowds. *Computer Graphics Forum 24*, 2 (2005), 203–215. 2

[RFWB07] RAMANARAYANAN G., FERWERDA J., WALTER B., BALA K.: Visual equivalence: towards a new standard for image fidelity. *ACM Trans. Graph. 26* (2007). 2

[RGK*08] RITSCHEL T., GROSCH T., KIM M. H., SEIDEL H.-P., DACHSBACHER C., KAUTZ J.: Imperfect shadow maps for efficient computation of indirect illumination. *ACM Trans. Graph. 27*, 5 (2008), 1–8. 2

[RH01] RAMAMOORTHI R., HANRAHAN P.: An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), SIGGRAPH '01, pp. 497–500. 1, 3

[SFWG04] STOKES W. A., FERWERDA J. A., WALTER B., GREENBERG D. P.: Perceptual illumination components: a new approach to efficient, high quality global illumination rendering. *ACM Trans. Graph. 23*, 3 (2004), 742–749. 2

[SGA*07] SUNDSTEDT V., GUTIERREZ D., ANSON O., BANTERLE F., CHALMERS A.: Perceptual rendering of participating media. *ACM Trans. Appl. Percept. 4*, 4 (2007), 15:1–15:22. 2

[SKS02] SLOAN P.-P., KAUTZ J., SNYDER J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Trans. Graph. 21* (2002), 527–536. 3

[VCF*09] VANGORP P., CONDON T. S., FERWERDA J. A., BALA K., SCHOUKENS R., DUTRÉ P.: *Visual Equivalence in Dynamic Scenes*. Tech. rep., Katholieke Universiteit Leuven, 2009. 2

[YCK*09] YU I., COX A., KIM M. H., RITSCHEL T., GROSCH T., DACHSBACHER C., KAUTZ J.: Perceptual influence of approximate visibility in indirect illumination. *ACM Trans. Appl. Percept. 6*, 4 (2009). 1, 2

[YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph. 20*, 1 (2001), 39–65. 2