Toward Evaluating the Usefulness of Global Illumination for Novices in Lighting Design Tasks

Ondřej Karlík, Martin Růžička, Václav Gassenbauer, Fabio Pellacini, and Jaroslav Křivánek

Abstract—Thanks to its ability to improve the realism of computer-generated imagery, the use of global illumination has recently become widespread among digital lighting artists. It remains unclear, though, what impact it has on the lighting design workflows, especially for novice users. In this paper we present a user study which investigates the use of global illumination, large area lights, and non-physical fill lights in lighting design tasks, where 26 novice subjects design lighting with these tools. The collected data suggest that global illumination is not significantly harder to control for novice users that direct illumination, and when given the possibility, most users opt to use it in their designs. The use of global illumination together with large area lights leads to simpler lighting setups with fewer non-physical fill lights. Interestingly, global illumination does not supersede fill lights: users still include them into their globally illuminated lighting setups. We believe that our results will find use in the development of lighting design tools for non-expert users.

Index Terms—Global Illumination, Lighting Design, User Study

1 Introduction

Lighting design is essential in defining the appearance of computer generated imagery [1]. While lighting in reality is confined to the laws of physics, computer artists have the luxury of being able to choose the appropriate *lighting model*, that is the mathematical model of light *emission* as well its *propagation* through the scene. Various lighting models have been developed, often trading off physical plausibility for computational efficiency. It is generally believed that physically plausible lighting models, global illumination in particular, facilitate the lighting design task and make it easier to create natural-looking images [2].

A common criticism of physically plausible lighting models in lighting design includes limited artistic freedom and long rendering times. Moreover, it is natural to expect that, due to its indirect character, global illumination should be more difficult to control than direct illumination. However, evidence is missing in the research community to support these beliefs. In addition, little is known about the influence of different lighting models on the lighting design workflow of *novice* users. In particular, will global illumination be beneficial for novice users, given the aforementioned advantages and drawbacks? Will they be able to control it and use it effectively? Answers to such questions are important for the development of lighting design tools targeted at non-expert users.

- *O. Karlík and J. Křivánek are with Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic. E-mails: karlik@cgg.mff.cuni.cz, jaroslav.krivanek@mff.cuni.cz.*
- *M. Růžička: E-mail: martinruzickatm@seznam.cz*
- *Václav Gassenbauer: E-mail: vgassenb@gmail.com*
- *F. Pellacini is with Dartmouth College and Sapienza University of Rome. E-mail: fabio@cs.dartmouth.edu*

In this paper we take the first steps in a quantitative evaluation of the impact of different lighting models on the performance of novice users in lighting design tasks. We focus on *global illumination* (GI), *large area vs. point light sources*, and *fill lights* (lights that do not cast shadows nor create specular highlights). Our work disregards the computational efficiency aspect of advanced lighting models, which is likely to be resolved in a not so distant future [3], and focuses solely on usability issues. We restrict our investigation to novices because they make up a potentially important user group and little is known about their behavior. Nowadays novice users can perform tasks, such as adjusting photos, editing video, etc., that were traditionally reserved for experts. It is likely they will be interested in creating images from 3D content, as soon as the latter becomes widespread. In addition, unlike expert users, novices do not have a priori preference with respect to different lighting models that could bias the study results.

In our study, 26 novice users perform various lighting design tasks in four scenes with different lighting models using a custom-built real-time relighting tool. Similar to Kerr and Pellacini [4] we ask subjects to finish two types of tasks. In *matching trials* they try to match lighting exactly using various lighting models. In *open trials*, which are more akin to the actual practical lighting design tasks, subjects set up lighting taking inspiration from a given image or according to their aesthetic preference. We collect data by recording all user actions and by asking subjects to fill out a questionnaire. By analyzing the data, we make the following conclusions.

1) *Novices can light with GI.* Their performance in matching trials is slightly worse with GI than with direct illumination, but it does not impair their

ability to finish the task. Users did not report increased task difficulty with GI in open trials.
2) *Novices want to light with GI.* When given the choice, most users opt to use GI in their designs.
3) *Novices create simpler lighting setups with physically plausible lighting models.* Users tend to use fewer lights in their designs when arbitrarily large lights and GI are allowed, provided that the lighting setup is complex enough.
4) *Fill lights remain useful.* Even if arbitrarily large lights and GI are allowed, users still employ some fill lights in their designs.

2 Related Work

In this section we briefly review the relevant work in fields of appearence design user studies, global illumination, and relighting engines.

2.1 Appearance design user studies

Kerr and Pellacini [4] evaluate the usability of different *user interfaces* in lighting design tasks. Our study differs significantly in that we focus on the impact of different *lighting models*, global illumination in particular, on the users' performance in such tasks. In another study, Kerr and Pellacini [5] evaluate the relative advantages of different user interfaces in material design tasks. Křivánek et al. [2] informally discuss some aspects of the use of GI in professional practice but their observations are not supported by experimental data. Reiner et al. [6] develop tools for visualizing light transport and show their usefulness in a number of specialized tasks. We focus on more traditional tools for lighting design.

2.2 Global illumination in lighting design

Global illumination [7] can provide remarkable realism when rendering synthetic scenes, and it has recently become a common tool in a number of applications, including architectural visualization, film production, or even video games [2], [8], [9]. One of the limitations of global illumination in artistic applications such as cinematic lighting is limited flexibility, constrained by the physical laws. Common approaches to artistic control of GI include rendering in layers and the use of compositing tools, or procedural modification of surface and light shaders [8], [10]. Obert et al. [11] and Schmidt et al. [12] introduce a representation for artistic control of global illumination. In our study we adhere to physically plausible global illumination without any artistic modifications. Doing so allows us to strictly define the different lighting models compared in the experiments.

2.3 Relighting engines

Lighting design is a trial-and-error process that involves testing many variants before deciding upon the final result. Relighting engines such as those described by Pellacini et al. [13] and Ragan-Kelley et al. [14] facilitate this process by providing the user with fast feedback. Though other alternatives exist (e.g. [15]), the implementation of the relighting engine used to run our study is based on the Direct-to-Indirect Transfer algorithm of Hašan et al. [16] because it most closely suits the needs of our application. Pellacini [17] proposes a lighting design interface for environment illumination; our study focuses on local lights only as they are more widely adopted in practice.

3 Goal and Motivation

We seek to evaluate the impact of different lighting model features on novice users' performance in lighting design tasks. Specifically, we want to answer the following questions:

1) What is the impact of global illumination on novice users' ability to precisely adjust lighting?
2) Are novice users able to control the indirect effects of global illumination (light bounced off of scene surfaces)?
3) What is the impact of the physically plausible lighting model features (namely large area lights and global illumination) on the complexity of the created lighting setups (i.e., number of lights)?
4) Does the use of these physically plausible features obviate the need for (non-physical) fill lights?
5) When free to choose, which lighting model features (GI, large area lights, fill lights) will novices use in their designs?

The first two questions are motivated by the surmise that global illumination, due to its indirect nature, could be more difficult to control than direct illumination, especially for novice users. This added difficulty could then counter the advantages that global illumination is believed to have in terms of the provided image realism.

An argument in favor of the use of GI in professional practice is a reduced lighting setup complexity and the consequent savings of lighters' effort [2]. Question (3) is motivated by the desire to quantitatively investigate this belief in the context of novice users. We also want to study the relative impact of physically plausible light *emission* (large area lights) and *transport* (GI).

Fill lights (non-physical lights that do not cast shadows nor create highlights) are commonly used in computer cinematography [1]. We ask Question (4) because we wish to find out if their popularity is a mere artifact of the inability of point lights to fill a scene with soft light. Or will they remain useful even when physically plausible soft light tools become available?

Answering Question (5) should reveal the relative advantages of the individual lighting model features as perceived by the novice users.

4 Experiments

Our study consists of experiments designed to provide answers to the above questions, where subjects are asked

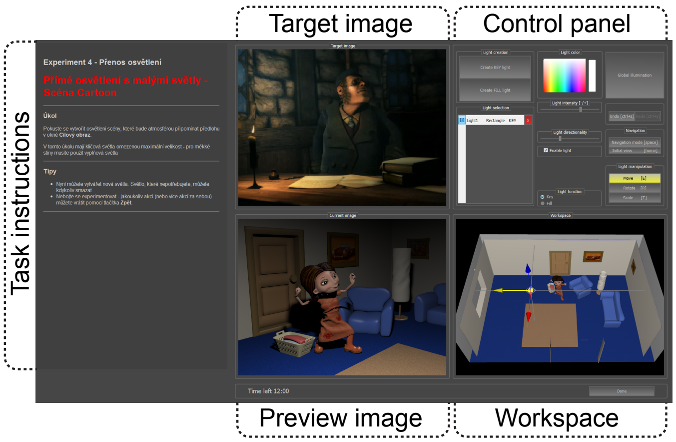

Fig. 1. User interface used to run the study.

Fig. 2. Scenes used in our study: Global view (top row) and the view used in the study (bottom row).

to perform various lighting design tasks with different lighting model variants.

4.1 Lighting model

Our base lighting model was designed to be as simple as possible but expressive enough to allow lighting design in realistic scenes. Since we study physically plausible illumination, we do not use point lights. Instead, all light sources are modeled as light-emitting squares (up to 3 degrees of freedom (DOFs) for position, 3 for orientation, and 1 for size) where radiance L_e emitted from $\mathbf{x}$ in direction ω is given by: $L_\mathrm{e}(\mathbf{x}, \omega) = L_0 \frac{d+2}{2\pi} \max\{0, \mathbf{n} \cdot \omega\}^d$. Radiance L_0 along the light normal $\mathbf{n}$ as well as the 'directionality' parameter d are specified by the user. We use the Modified Phong BRDF for surface reflectance [18, p. 32] and we always apply the physical inverse-squared-distance intensity falloff when computing direct illumination.

Our definition of *key* and *fill* lights follows the standard practice [1], [4]. *Key lights* create highlights on surfaces and cast shadows, while there is no visibility checking for *fill lights* (no shadows), and only the diffuse BRDF component is taken into account for them (no specular highlights).

High-quality images free of visual artifacts of any kind are crucial to avoid bias in user preferences. For that reason we refrain from using shadow mapping or interpolated indirect illumination. Instead, we use accurate per-pixel calculations that deliver images at interactive rates. Direct illumination with soft shadows is computed using quasi-Monte Carlo integration implemented in CUDA [19]. The global illumination variant of our lighting model adds an *indirect* component as per the standard path tracing expansion of the rendering equation [7]. We compute it using a modified version of the direct-to-indirect transfer algorithm [16] (please see the supplemental document for details).

There are two limitations of the rendering system used. First, since it takes about a second to compute the high-quality direct lighting, lower number of samples is used when lights are being manipulated to keep the interaction real-time (i.e., there is visible Monte Carlo noise during interaction when using large area lights). Immediately after the manipulation ends, noise-free image is computed and displayed. This behavior is shown in the supplemental video; users had no problems with it during the experiments. Second, due to the particularities of the direct-to-indirect transfer algorithm, the first bounce from the light source is restricted to Lambertian reflection, so the the indirect illumination corresponds to paths of type $LD\{S|D\}^*E$ in Heckbert's notation [7]. Note that these paths *do include specular interreflections*.

4.2 User interface

Our choice of the light manipulation user interface follows the results of Kerr and Pellacini [4]. Because the indirect light manipulation (shadow or highlight dragging) is not applicable to the smooth illumination we focus on, we use the *direct light manipulation* paradigm with 'gizmos' implemented in a similar way to *Maya* [20]. The user interface, shown in Fig. 1, consists of the *task instruction panel*, *target image*, *preview image* (fixed camera rendering of the scene illuminated by the current lighting setup), *workspace* (OpenGL rendering of the scene with simple constant lighting, where users can navigate and manipulate lights), and *control panel*. The controls include: light creation buttons, light selection list, light attribute controls (on/off button, color, intensity, directionality, key/fill light switch), GI on/off button, undo/redo buttons. Most trials require only some of these controls; we always hide the rest of them. We use a fixed camera view for the preview image because it is a common practice in cinematic lighting design. Furthermore, an arbitrary viewpoint would unjustifiably complicate the work and also raise the question of how to meaningfully measure matching error. Thorough pilot tests have shown that novice users can comfortably use this interface after a short training.

4.3 Scenes

Fig. 2 shows the four scenes used in our study. We specifically choose scenes that serve different purposes

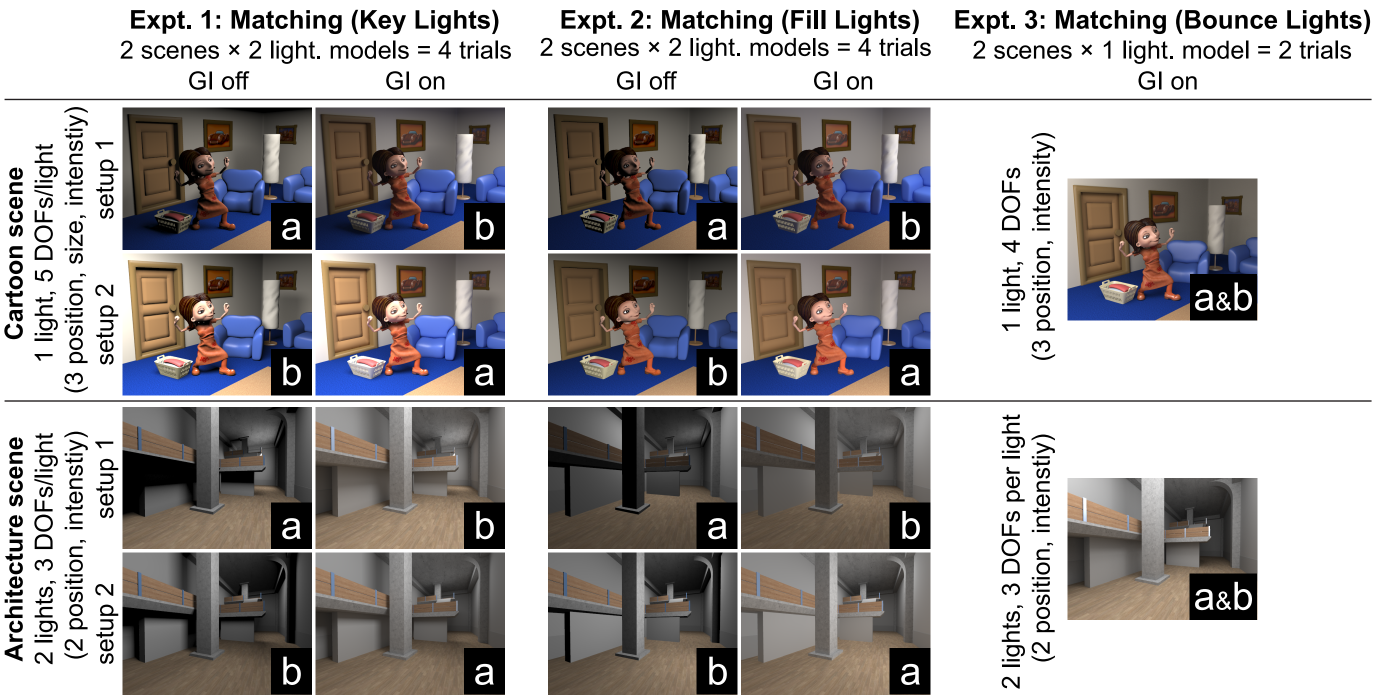

Fig. 3. Target images for matching Experiments 1 to 3. Subjects are asked to exactly match the target image. They work with the same scene (Cartoon, Architecture) and lighting model (global illumination off/on) as was used to create the target image. The experiments differ by the type of light that the subjects adjust: key, fill, and bounce (key lights bounced off ceiling or walls), respectively. Each user matches only half of the images in Experiments 1 and 2 (either those marked “a” or those marked “b”), and both shown images in Experiment 3.

(entertainment, architectural visualization), exhibit different geometric setting and complexity, and are affected by global illumination to a different degree (Still life vs. other scenes). We did not include scenes with highly glossy light transport effects (e.g. caustics) because of the limitations of the relighting algorithm, and because we believe that the precise matching of such localized effects is orthogonal to the questions we set to answer. We did not ask users to match images with sharp shadows for the same reason. We believe that our scenes are representative of what a user might encounter in practice.

4.4 Trials

The study is made up of five experiments summarized in Fig. 3 and 4. Experiments 1 to 3 consist of *matching trials*, where the subjects are asked to match the given target image by adjusting parameters of a fixed number of lights. The target images for these trials were created using the same camera view, same type and number of lights, as well as the same lighting model variant that the users have at their disposal, and therefore *can be matched exactly*. The matching trials allow us to quantitatively measure subjects’ performance, while providing them with a clear goal.

The purpose of **Experiments 1 and 2** (see Fig. 3) is to investigate the impact of global illumination on the users’ ability to precisely adjust lighting, either in the presence of distinct visual cues like shadows and highlights (Experiment 1) or when no such cues are present (Experiment 2). In Experiment 1 subjects manipulate *key lights* while in Experiment 2 they manipulate *fill lights*. We investigate two lighting model variants (global illumination off/on) in two scenes (Cartoon, Architecture), giving a total of *4 trials per subject*. To avoid learning effects, one subject is never asked to match two target images created using the same lighting setup, that only differ in the lighting model. For that reason, a subject matches two different target lighting setups for each scene, one with GI on and the other with GI off. Users work in each trial with a fixed number of lights, placed initially near the center of the scene. The number of degrees of freedom (5 for the Cartoon and 6 for the Architecture scene, see Fig. 3) was calibrated based on pilot studies such that most users could finish the tasks.

In **Experiment 3** we investigate the users’ ability to control the indirect effects of global illumination by asking them to match the target image using indirect (bounced) lighting. Users manipulate key lights that are constrained in size (Cartoon) and also in position (Architecture) so that they are forced to use a wall or the ceiling as a bounce plane and illuminate the visible part of the scene indirectly. Each subject does one trial per scene (*2 trials total*). Note the important difference from Experiments 1 and 2 where we measure the impact of the presence or absence of GI on the ability to match (predominantly direct) lighting.

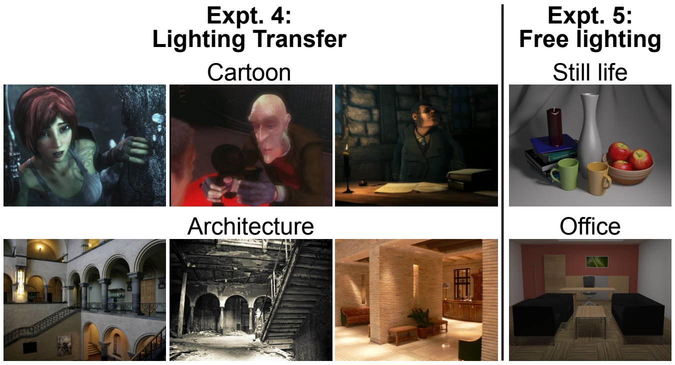

Fig. 4. Left: Target images used in Expt. 4. Subjects are asked to set up lighting following inspiration from these target images. Two scenes with 3 different lighting models are used, resulting in 6 tasks total. Right: Scenes used in Experiment 5. Subjects are asked to set up lighting in these scenes according to their aesthetic preference. Subjects do two tasks, each one in a different scene.

Experiments 4 and 5 are *open ended* in that there is no 'right' solution. Subjects are asked to design lighting without a target image that could be matched exactly. Such tasks are more akin to lighting design in practice than matching, but do not allow to objectively measure accuracy.

The purpose of **Experiment 4**, which we refer to as *lighting transfer*, is to find out how the a) lighting setup complexity (number of lights) and b) the use of fill lights depend on the lighting model. Subjects set up lighting in the Cartoon and Architecture scenes drawing inspiration from target images (see Fig. 4). To aid subjects in performing this task, the target images have a distinctive stylized lighting and roughly match our scenes (a frame from a cartoon animation, and an architectural scene).

In this experiment users work with three lighting models:

- LM1 direct illumination with key lights limited to small size (with sharp shadows)
- LM2 direct illumination with key lights of arbitrary size, and
- LM3 same as LM2 but with GI on,

in two scenes, yielding a total of *6 trials per subject*. The light size restriction in LM1 only applies to key lights; fill lights can always have arbitrary size. Subjects can create any number of lights – the number of DOFs is not limited. The assignment of target images to lighting model is randomized. The above lighting models allow us to measure the effect of physically plausible light *emission* (LM2) and *transport* (LM3).

In **Experiment 5** (*free lighting*), we investigate users' preference in terms of lighting model features when given a choice. Subjects are asked to design lighting in the Still life and Office scenes according to their aesthetic preference without any target image (see Fig. 4). We switch to different scenes in Experiment 5 to prevent subjects from merely repeating a previous design they have created. The lighting model is completely flexible, users can create any number of (arbitrarily sized) lights and turn GI on and off at any time. They are specifically instructed to experiment with the latter feature.

To alleviate learning effects we randomize the task order as well as the target image / lighting model assignment for each subject in all experiments. The supplemental video includes an example user workflow for each of the five experiments.

4.5 Questionnaire

Subjects fill out a questionnaire as they progress through the study. In Experiments 1 to 3, they are asked to rate, immediately after each trial, a) subjective match quality and b) task difficulty. In Experiment 4, they rate a) subjective image quality, b) task difficulty, and c) 'restrictiveness' (i.e., to what degree they felt restricted by the lighting model in expressing their design ideas). In Experiment 5 they rate and rank the usefulness of the following features for lighting design: a) arbitrarily sized key lights, b) fill lights, and c) global illumination. All ratings are on a 1 to 5 scale. Immediately after finishing the trials for one experiment, subjects are asked to leave free-form comments on their workflow and ratings. The filled-out questionnaires are included as supplemental material

4.6 Procedure

Twenty six paid subjects chosen from different gender, age and educational groups, all with normal or corrected-to-normal vision, participated in the study. They had no or little previous experience with digital lighting. The study consists of four parts (training, Expts. 1 to 3, Expt. 4, and Expt. 5) separated by breaks. Subjects are free to take additional breaks. In the *training* part (40–60 min), subjects are acquainted with the purpose of the experiments and instructed on all aspects of the interface through a series of elementary tasks. We train subjects individually to allow for questions and to accommodate each subject's learning needs. Details of this procedure are in the supplemental material. The instructor verifies that the subject understands the task, and answers her/his questions. Once the experiment begins, all user interface actions are recorded. Before each experiment the instructor explains its particularities. To constrain the study duration, there is a time limit for each trial (8 min in Experiments 1 to 3 and 15 min in Experiments 4 and 5). Most subjects finished the entire session in 4 to 5 hours.

The study was conducted in a lighting environment with low ambient lighting to simulate typical working conditions of artists and maximize visibility of the screen. We used a Dell U2711 27″ monitor with a native resolution of 2560×1440 at approximately 60 cm from the subject. All images are rendered at 640×480 pixels and tone mapped with a fixed exposure-gamma algorithm ($\gamma = 2.2$).

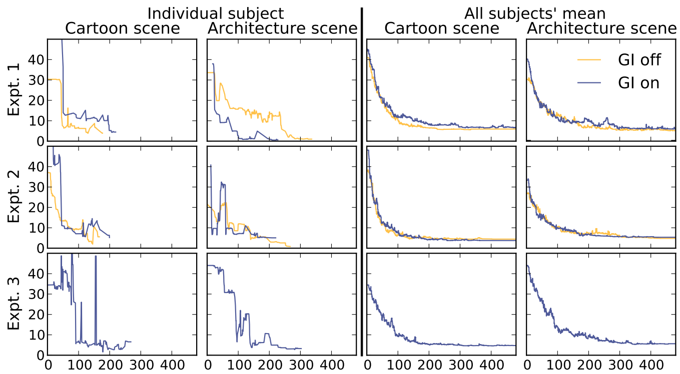

Fig. 5. CIELAB error plotted over time for the matching trials for an individual subject (left) and averaged across all 26 subjects (right). The graphs are shown separately for the Cartoon and Architecture scenes.

Before the actual study we run the experiments with multiple pilot subjects to ensure that users can comfortably use the interface, finish the tasks within time limits, and to make sure that the application speed and output quality is adequate for them.

5 Analysis

In this section we present and analyze the study results. Following the advice of many authors (e.g. [21], [22]), we use *confidence intervals* rather than the traditional significance tests. This is because significance testing only expresses the confidence that the population means of two groups differ (e.g., that GI has an effect on user performance), but gives no information about the magnitude of the difference (*effect size*). The effect size can be estimated directly using indicators such as *Cohen's d* or *Pearson's r* [23]. These, however, in turn give no information on the confidence in the results.

Confidence intervals (CIs), on the other hand, express information about both the effect size and confidence in the result. An $\alpha\%$ confidence interval is an interval estimate of the unknown population parameter (e.g., increase in task difficulty when using GI), constructed from measured data. If we were to repeat the experiment many times, we would measure different data every time, and therefore construct different confidence intervals. Then, $\alpha\%$ of these intervals would include the true population parameter. Note that α *is not the probability, that any particular confidence interval contains the population parameter.*[1] Confidence intervals are closely related to significance testing – the latter judges a difference of two populations significant with $p = 1 - \alpha$ if and only if the α-confidence interval of the difference does not contain zero [24]. This means that confidence intervals show a superset of information provided by significance testing.

1. Such intervals exist, and are called *credible intervals*. Their construction, however, would require substantial a priori knowledge about the data.

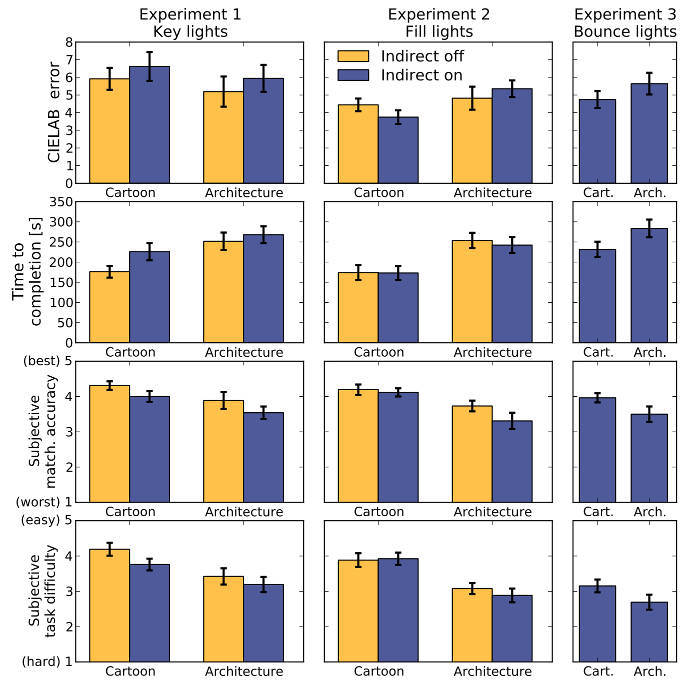

Fig. 6. Objective and subjective quantities measured in Experiments 1 to 3 averaged across all subjects. Top to bottom: final matching error, time to completion, subjective match accuracy, subjective task difficulty.

We report the 90% confidence intervals for inter-subject differences (e.g. GI off–GI on) calculated using the paired t-test [22] where appropriate. Details are given in the supplemental material. To facilitate comparison between confidence intervals of quantities with different domains, we divide them with the mean value of the first group and report the intervals in percent. Note that this transformation does not alter the conclusions in any way. We report the confidence intervals as $mean\%$ ($[low,\ high]\%$), where low and $high$ are the bounds of the confidence interval, and $mean$ is the measured difference between groups, which is also the center of the interval. We also report sample means for the individual trials using bar graphs together with the standard error of each mean, drawn as a black error bar. Their numerical values, as well as more indicators computed from the data (namely 95% CIs, Cohen's *d*, Pearson's *r*, and results of traditional significance testing), are provided in the supplemental material.

5.1 Matching Trials (Experiments 1 to 3)

The first three experiments consist of matching trials. Subjects are asked to match a target image as closely and quickly as possible using key lights (Experiment 1), fill lights (Experiment 2), or an indirect bounce of key lights (Experiment 3). We measure objective image error, time to completion, and subjective ratings of match accuracy and task difficulty provided by the users in the questionnaire.

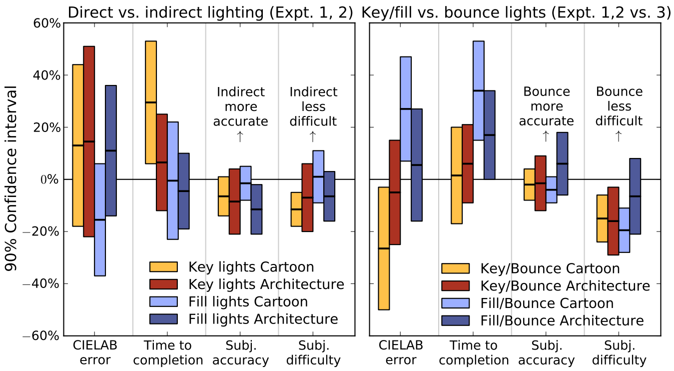

Fig. 7. 90% confidence intervals for our statistical analysis of Experiments 1 to 3. Each bar is an interval estimate of the effect size of either turning on GI (left), or using bounce instead of key/fill lighting (right). Effect sizes for different measured quantities (different groups in the figure) and for different scenes and light types (different colors) are shown. Center of a bar further from zero means bigger effect size; shorter bar length indicates bigger confidence in it. A result is statistically significant if the bar does not overlap the 0% line.

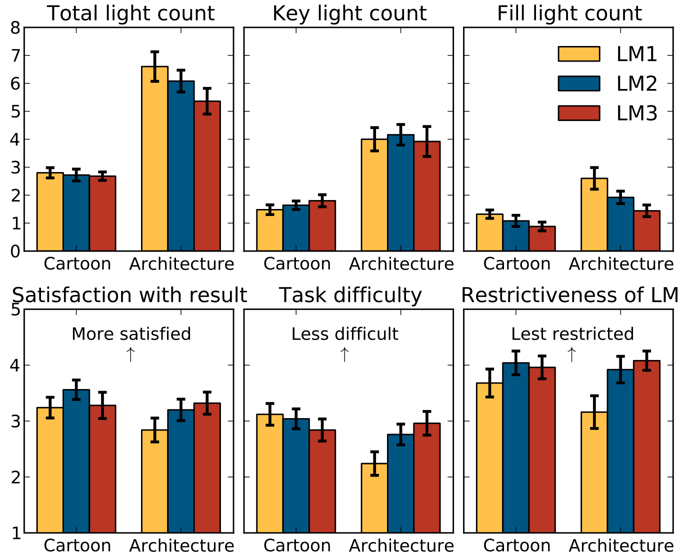

Fig. 8. Results of Expt. 4. Top row: number of lights in users' designs. Bottom row: subjective measurements (satisfaction with the result, task difficulty, lighting model restrictiveness). The lighting models are: LM1: small key lights with GI off, LM2: arbitrarily large key lights with GI off, and LM3: same as LM2 with GI on.

5.1.1 Matching error

We measure the image error as the Euclidean distance in the perceptually unifrom CIELAB color space. For two images A and B it is computed as

$$\frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(L_{A,i}-L_{B,i}\right)^2+\left(a_{A,i}-a_{B,i}\right)^2+\left(b_{A,i}-b_{B,i}\right)^2},$$

where $n = 640 \times 480$ is the number of pixels and $L_{X,i}, a_{X,i}, b_{X,i}$ are the values of L, a, b channels of i-th pixel of image X (either A or B). We opt to use this simple per-pixel metric because more complicated visual difference metrics have been shown to sometimes provide unreliable results [25]. Our choice of metric is also further validated by the high correlation of its outputs with subjective user satisfaction with the result.

We keep track of the image error during the entire course of each trial. The error plots in Fig. 5 show an overall decrease of the error over time, indicating that the subjects are able to perform the tasks in all three experiments irrespective of whether global illumination is on or off. Similar graphs for all subjects and trials are available in the supplemental material.

The objective matching accuracy is measured using the *final error*, which is the image error present when users end the task. The mean errors are shown in Fig. 6, confidence intervals for their differences in Fig. 7. The error is higher for global illumination in 3 out of 4 paired (i.e. GI on-off) trials in Experiments 1 and 2. The biggest difference is in key light matching in the Architecture scene, where enabling GI increases error by 14% (CI: $[-22,\ 51]\%$). The GI has opposite effect while matching fill lights in the Cartoon scene – its presence decreases the error by 16% ($[-6,\ 37]\%$). Overall, it seems that global illumination might slightly increase the final error when precisely matching lighting, though the result is not conclusive.

Next, we compare the final errors in Experiments 1 and 2 (key/fill lights) with final errors in Experiment 3 (lighting bounced off walls). In the Cartoon scene using bounce instead of fill lights increases the error by 27% ($[7,\ 47]\%$), but using bounce lights compared to key lights decreases the error by 27%, $[3,\ 50]\%$. There are no such strong differences in the Architecture scene.

5.1.2 Time to completion

The time the subjects took to finish the trials indicates the efficiency of users' work with different lighting models. We show the mean time for each trial in Fig. 6 and the confidence intervals in Fig. 7. The time limit of 8 minutes for all matching trials was sufficient in most cases; a trial was terminated because of it only 8 times out of 260. Global illumination has strong effect only in the Cartoon scene for key light matching, where it increases the time needed by 29% ($[6,\ 53]\%$). In other cases it has insignificant effect (CIs lying in $[-23,\ 25]\%$). Matching with bounce lights took users slightly more time compared to fill lights, i.e. 34% ($[15,\ 53]\%$) difference in the Cartoon scene, 17% ($[0,\ 34]\%$) for Architecture. The effect size for key vs. bounce lights was negligible, with both 90% CIs lying in $[-17,\ 21]\%$.

5.1.3 Subjective ratings

Users rated subjective accuracy of their matches as well as the task difficulty for each trial on a scale from 1 to 5

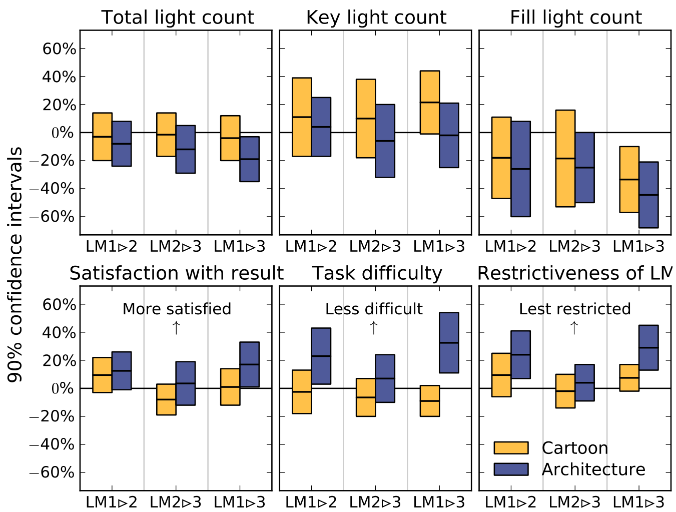

Fig. 9. 90% confidence intervals for results of Experiment 4. The meaning of the bars is the same as in Fig. 7. LM1▷2 is the transition from LM1 to LM 2, which means enabling arbitrary size key lights. Similarly, LM2▷3 means turning GI on.

(see Fig. 6 and 7). *Subjective accuracy* results confirm the observed trend of the objective matching error. In all 4 paired trials in Expts. 1 and 2 the mean accuracy ratings are lower with GI on than with GI off, though the differences are very small – all CIs lie between $[-21,\ 5]\%$. *Subjective task difficulty* shows similar trends. Users generally rate lighting with GI as more difficult, with the biggest difference being in matching key lights in Cartoon scene (-11%; $[-18,\ -5]\%$).

Interestingly, users rate matching bounce lights consistently harder than key/fill lights. The biggest difference was observed when matching fill lights in the Cartoon scene: 20% ($[11,\ 28]\%$); smallest when matching fill lights in the Architecture scene: 7% ($[-8,\ 21]\%$). Despite the increased difficulty, the subjective differences in accuracy were small, with all CIs bounded by $[-12,\ 18]\%$.

5.1.4 Workflow

Most users listed the same basic steps of their workflow: first adjusting light intensity to roughly match overall scene brightness, then locating prominent lighting features (such as shadows, highlights, etc.) and matching them by manipulation with light position, and finally fine-tuning the light intensity. This workflow was confirmed by inspecting the recorded videos and by questionnaires. Most users do not admit any major differences in their workflow based on lighting model (i.e. GI setting) or scene, aside from using different cues for estimating light position (e.g., shadows for matching with key lights and bright areas for fill lights). Additionally, the only noticeable difference in workflows between bounced light matching (Expt. 3) and rest is that the former is slightly more trial-and-error based.

5.2 Open Trials (Experiment 4 – Lighting Transfer)

In Expt. 4 subjects design lighting taking inspiration from a target image only vaguely related to the scene they work with. They use 3 lighting models (LM1: GI off, small key lights, LM2: GI off, arbitrarily large key lights, LM3: GI on, arbitrarily large key lights). Some of the result images from this experiment are shown in Fig. 10. As this is an open ended task with no 'right' solution, there is no objective error metric. Instead we measure the number of lights users create to see how the lighting model influences the lighting setup complexity, and collect subjective ratings of satisfaction with the result, task difficulty, and restrictiveness of the lighting model.

5.2.1 Number of lights used

We record the number of active key and fill lights at the end of each trial. The data, shown in Fig. 8, have rather high variance, but if we look at confidence intervals in Fig. 9, we can see a strong trend of decreasing number of fill lights as the lighting model gets more physically plausible. Moving from LM1 to LM3 decreases the number of fill lights used by 33% (CI: $[10,\ 57]\%$) for the Cartoon and by 45% ($[21,\ 68]\%$) for the Architecture scene. The number of key lights shows opposite trend in the Cartoon scene – moving from LM1 to LM3 increases their count by 22% ($[-1,\ 44]\%$). The same transition has no clear effect in Architecture scene. Overall, moving towards more physically plausible light model strongly decreases the number of fill lights needed, but does not have a clear effect on the number of key lights.

5.2.2 Time to completion

The meaning of this quantity is not as well-defined as in the matching trials due to the higher subjectivity of lighting transfer trials, but we include it in the supplemental material for the sake of completeness. The time limit of 15 minutes was reached only in 2 trials out of 156. There is no significant dependence of time to completion on the lighting model.

5.2.3 Subjective ratings

The average subjective ratings provided by the users after each trial are shown in Fig. 8; the associated confidence intervals are in Fig. 9. The Architecture scene shows a clear trend of increasing satisfaction and decreasing difficulty and restrictiveness when moving towards more physically plausible lighting model: Between LM1 and LM3 we measured increase in satisfaction by 17% ($[1,\ 33]\%$), decrease in task difficulty by 32% ($[11,\ 54]\%$) and decrease of lighting model restrictiveness by 29% ($[13,\ 45]\%$). There is no definitive trend in the subjective rating in the Cartoon scene (all CIs are between $[-20,\ 17]\%$). We believe that this might be because the simpler nature of the scene allows users to light it reasonably well even with a less powerful toolset.

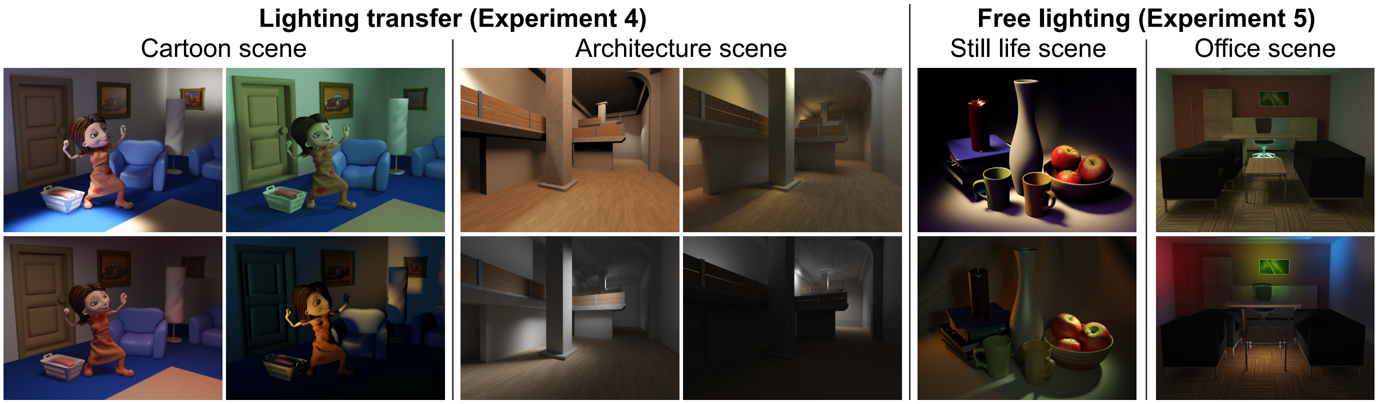

Fig. 10. Examples of test subjects' final lighting setups in the open-ended experiments 4 and 5. Novice users were able to achieve these results in under 15 minutes. Final images for all subjects are available in the supplemental material.

The differences in the Architecture scene are explained in the questionnaire: users often wanted to use the features that were disabled in LM1 and LM2, and tried to simulate them with tools at hand, which influenced the difficulty of the task: "It (difficulty) depends on the difficulty of replacing disabled lighting features.", "I was surprised how hard it is to replace a large key light".

5.2.4 Workflow

The workflow in this experiment was much more varied. Most users tend to start from the most prominent lighting features (visible lights, hard shadows, etc.) and then fine-tune the atmosphere of the image. Overall the workflow is more trial-and-error based than for matching. The only observed difference between lighting workflow with and without GI is that subjects used more fill lights without it, which, in addition to being objectively measured, was indicated by the users themselves in the questionnaire, e.g: "Indirect lighting looks more natural, so when it was turned off, I used a large number of weak fill lights to replace it."

5.3 Open Trials (Experiment 5 – Free Lighting)

In Experiment 5 users design lighting according to their aesthetic preference with no target image given. The lighting model is completely flexible, and the users are free to turn GI on and off. Some of the subjects' results are shown in Fig. 10. We analyze user preferences of the following lighting model features: 1) arbitrarily large key lights, 2) fill lights, and 3) global illumination, as expressed via questionnaire and via their actual use in the created lighting setups.

Note that in this section we also report *proportions* of users exhibiting certain traits (e.g. using GI). Because of the non-normality of this data, we have to construct the confidence intervals using different method, namely Wilson score intervals [26]. The semantics of proportion CIs are unchanged, with the only difference being that they are not symmetrical around the measured mean.

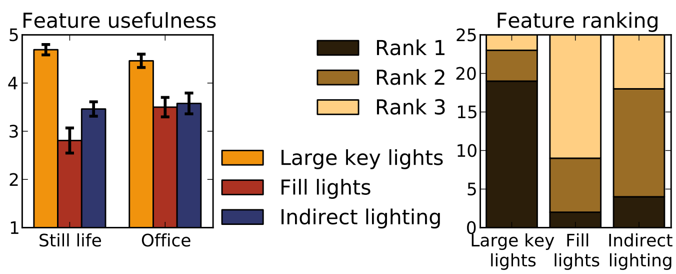

Fig. 11. Subjective ratings (left) and rankings (right) of the usefulness of lighting model features for the free lighting tasks (Expt. 5). Only 25 rankings are included as one subject provided invalid data.

5.3.1 Lighting model features used

Users had the freedom to turn GI on and off at any time. Out of 26 subjects, 16 left it on in their final setup in the Still life scene and 20 in the Office scene. Most subjects also decided to use fill lights; 16 used at least one in the Still life and 20 did the same in the Office scene. The 90% confidence interval for population proportion that opted to use GI is $[45,\ 75]\%$ in the Still life scene and $[61,\ 88]\%$ in the Office scene. The confidence intervals for fill light usage are the same ($[45,\ 75]\%$ in Still life scene, $[61,\ 88]\%$ in Office scene). This is just a coincidence – even though the proportions are the same, there was no significant correlation between these two choices. Overall, we have observed that the majority of subjects use both GI and fill lights. Additionally, more subjects use these advanced features in the Office scene, indicating that the scene may be more challenging to light.

5.3.2 Workflow

Users tend to decide whether or not to use GI early on, and generally do not change the decision later. Many of them affirm the choice later by switching between GI on and off, comparing the results, and returning back to the original decision.

5.3.3 Subjective rating of features

We have gathered users' opinions on the usefulness of different lighting model features in two ways: by having them *rate* the usefulness in both trials on a scale from 1 to 5, and *rank* the features by overall usefulness after all trials. The results are summarized in Fig. 11. The ratings show a strong preference for arbitrarily large key lights, that are rated better than indirect lighting by 26% (CI: $[19,\ 34]$%) in the Still life and by 20% ($[9,\ 31]$%) in the Office scene. Additionally, GI is rated better than fill lights in the Still life scene by 23% ($[5,\ 41]$%); the same comparison in the Office scene is inconclusive (CI: $[-12,\ 17]$%).

5.3.4 Subjective ranking

By forcing subjects to choose, we get similar, but much stronger trends. 20 out of 25 valid answers ranked arbitrarily large key lights as more useful than GI (90% CI of population proportion: $[64,\ 90]$%), and 17 out of 25 ranked GI as more useful than fill lights (CI: $[51,\ 80]$%). This indicates, that most users consider soft shadows the most useful feature, and that most users would prefer GI to fill lights as a sole method for creating soft lighting (even though most acutally prefer to use both).

Various reasons for the ratings and rankings are mentioned in the questionnaire. Users praise the realism of GI ("Indirect lighting causes objects to look believable and organic."), and criticize fill lights for creating flat, unrealistic lighting ("Images with fill lights look quite flat, without any distinctive or interesting parts.") Despite the low rating, most people still use fill lights. The questionnaires suggest that users do not consider them crucial for the lighting, but use them as an auxiliary feature to help fine-tune the dark parts of the image (e.g., "I've used fill lights, but only as a supplement, main lights were always key lights."). This shows that even without specific training, the novices' use of fill lights is in line with the recommended professional practice [1].

6 Summary of Results and Discussion

In this section we summarize the most important results of our study, and discuss the limitations. We start by formulating answers to the questions defined in Section 3. We remind the reader of an intrinsic limitation of this (as well as any other) user study: strictly speaking, the results we have obtained apply only to our test cases.

6.1 Impact of GI on the ability to adjust lighting (Q1)

Contrary to our expectation, the presence of GI has relatively small effect on the subjects' performance in matching trials (as demonstrated by Experiments 1 to 3). While users generally perform better without global illumination, the effect size is small and often below the threshold of statistical significance.

6.2 Ability to light with indirect illumination (Q2)

After comparing questionnaire results of Experiments 1 and 2 to Experiment 3, we came to the conclusion that users find indirect (bounce) lighting more difficult to work with than direct lighting. Nonetheless, the measured data show that their ability to complete the task is not significantly affected. This, together with the observation made in the previous paragraph, and results in Fig. 5, suggests that while they may find it harder, *novices can light with GI*.

6.3 Lighting setup complexity (Q3)

When lighting with GI and large key lights, novice users tend to *use fewer fill lights* in their designs, provided that the scene is complex enough, as was shown in the Architecture scene in Experiment 4. The same behavior has been informally reported for expert users [2] and our result confirms it for novices. Surprisingly, we did not observe any corresponding reduction in time to completion.

6.4 Fill lights (Q4)

Even though non-physical fill lights are a concept unknown in the real world, according to Experiment 5, novice users are able to understand them and use them effectively. They use fill lights even with global illumination and large key lights. This suggests that *fill lights remain useful even with advanced physically plausible lighting models*.

6.5 Lighting model feature choice (Q5)

Novices choose to employ global illumination, large key lights, as well as non-physical fill lights in their designs, as confirmed by Experiment 5. For global illumination this indicates that the small increase of complexity of lighting with GI observed in the matching trials (Q1 and Q2) does not counter its advantage in terms of increased image realism.

6.6 Novices and complex lighting models

Even though most of our subjects had no previous lighting design skills, they were still able to effectively complete design tasks in realistic scenes using complex, physically plausible lighting models. This important finding generalizes a similar observation made by Kerr and Pellacini [4] in the context of lighting design in simple scenes using point lights.

6.7 Workflow

We found no significant impact of global illumination on users workflows in matching trials. In open-ended trials users were forced to create more fill lights to compensate for the unavailability of large key lights and/or global illumination. When free to choose whether or not to use GI, users make the decision at the beginning of the trial and usually do not change it later. In other words, when used, GI was a part of the lighting design process, not just a "beautify" add-on used after the lighting has been designed.

6.8 Limitations

Similarly to other user studies, our results are limited by the scope of our experiments, especially in terms of the lighting design tasks performed and the range of scenes that appeared in the study. Furthermore, while some of our results regarding the impact of GI on lighting design are in line with what has previously been reported for experts [2], it is unclear if the study results generalize to more advanced users. Finally, due to the technical limitations of our relighting engine, the images with global illumination are missing one particular type of light paths, which sometimes results in slight bias.

7 Conclusion and Future Work

This paper investigates the usefulness of advanced lighting model features, especially global illumination, for novice users. Through a user study conducted on 26 subjects, consisting of series of matching and open-ended trials, we have measured how a physically plausible lighting model features affect users' workflow.

The main result of the matching trials is that the presence of indirect illumination does not significantly worsen the results, which indicates that users do not have problems with its controllability or predictability. Open-ended trials show that novices prefer to use global illumination and large key lights when given the possibility. Its use also leads to simpler lighting setups. One additional interesting result is that even though *fill lights* are rated as the least useful feature, most novices still use them to fine-tune image details, regardless of whether they use global illumination or not. We believe that these results have immediate practical utility since they provide guidelines on the choice of a suitable lighting model in a lighting design application targeted at novice users.

We acknowledge that our measurements are in the strictest sense valid only for our test cases. Additionally, since the novices are still on the *learning curve*, we do not know whether they would change their preferences as they progress. This is an important open question to be resolved. Other possible direction of future research is broadening the *scope* of results – by including more scenes, lighting models, and tasks, or by including professional users. The mere fact, that novice users enjoy working on lighting design when provided with an intuitive user interface and high-quality interactive feedback, should also stimulate more research on truly real-time global illumination algorithms.

Acknowledgements

Jaroslav Křivánek and Ondřej Karlík were supported by the Marie Curie fellowship PIOF-GA-2008-221716 and the Czech Science Foundation grant P202-13-26189S. Fabio Pellacini was supported by the NSF (CNS-070820, CCF-0746117), Intel, and the Sloan Foundation.

We would also like to thank Jiří Friml for the Office scene.

References

[1] J. Birn, *Digital Lighting and Rendering*, 2nd ed. New Riders Press, 2006.

[2] J. Křivánek, M. Fajardo, P. H. Christensen, E. Tabellion, M. Bunnell, D. Larsson, and A. Kaplanyan, "Global illumination across industries," in *ACM SIGGRAPH 2010 Courses*. New York, NY, USA: ACM, 2010.

[3] C. Dachsbacher and J. Kautz, "Real-time global illumination for dynamic scenes," in *ACM SIGGRAPH 2009 Courses*. New York, NY, USA: ACM, 2009, pp. 19:1–19:217.

[4] W. B. Kerr and F. Pellacini, "Toward evaluating lighting design interface paradigms for novice users," *ACM Trans. Graph. (SIGGRAPH 2009)*, vol. 28, no. 3, pp. 26:1–26:9, 2009.

[5] ——, "Toward evaluating material design interface paradigms for novice users," *ACM Trans. Graph. (SIGGRAPH 2010)*, vol. 29, no. 4, pp. 35:1–35:10, 2010.

[6] T. Reiner, A. Kaplanyan, M. Reinhard, and C. Dachsbacher, "Selective inspection and interactive visualization of light transport in virtual scenes," *Computer Graphics Forum (EUROGRAPHICS 2012)*, vol. 31, no. 1, pp. 160–188, 2012.

[7] P. Dutré, K. Bala, and P. Bekaert, *Advanced Global Illumination*, 2nd ed. A K Peters, 2006.

[8] E. Tabellion and A. Lamorlette, "An approximate global illumination system for computer generated films," *ACM Trans. Graph. (SIGGRAPH 2004)*, vol. 23, no. 3, pp. 469–476, 2004.

[9] P. H. Christensen, "Point-based approximate color bleeding," Pixar, Emeryville, California, Tech. Rep. #08-01, July 2008.

[10] P. Christensen, "Global illumination and all that," in *ACM SIGGRAPH Courses (Course 9 – RenderMan, Theory and Practice)*. New York, NY, USA: ACM, 2003.

[11] J. Obert, J. Křivánek, F. Pellacini, D. Sýkora, and S. Pattanaik, "iCheat: A representation for artistic control of indirect cinematic lighting," *Computer Graphics Forum (EGSR 2008)*, vol. 27, no. 4, pp. 1217–1223, 2008.

[12] T.-W. Schmidt, J. Novak, J. Meng, A. S. Kaplanyan, T. Reiner, D. Nowrouzezahrai, and C. Dachsbacher, "Path-space manipulation of physically-based light transport," *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2013)*, vol. 32, no. 4, Aug. 2013.

[13] F. Pellacini, K. Vidimče, A. Lefohn, A. Mohr, M. Leone, and J. Warren, "Lpics: a hybrid hardware-accelerated relighting engine for computer cinematography," *ACM Trans. Graph. (SIGGRAPH 2005)*, vol. 24, no. 3, pp. 464–470, 2005.

[14] J. Ragan-Kelley, C. Kilpatrick, B. W. Smith, D. Epps, P. Green, C. Hery, and F. Durand, "The lightspeed automatic interactive lighting preview system," *ACM Trans. Graph. (SIGGRAPH 2007)*, vol. 26, no. 3, 2007.

[15] J. Lehtinen, M. Zwicker, E. Turquin, J. Kontkanen, F. Durand, F. X. Sillion, and T. Aila, "A meshless hierarchical representation for light transport," *ACM Trans. Graph. (SIGGRAPH 2008)*, vol. 27, no. 3, pp. 37:1–37:9, Aug. 2008.

[16] M. Hašan, F. Pellacini, and K. Bala, "Direct-to-indirect transfer for cinematic relighting," *ACM Trans. Graph. (SIGGRAPH 2006)*, vol. 25, no. 3, pp. 1089–1097, 2006.

[17] F. Pellacini, "envyLight: an interface for editing natural illumination," *ACM Trans. Graph. (SIGGRAPH 2010)*, vol. 29, no. 4, 2010.

[18] P. Dutré, "Global illumination compendium," 2003.

[19] T. Aila and S. Laine, "Understanding the efficiency of ray traversal on GPUs," in *Proc. High-Performance Graphics 2009*, 2009, pp. 145–149.

[20] Autodesk Inc, "Maya 2011," 2011.

[21] D. H. Johnson, "The Insignificance of Statistical Significance Testing," *The Journal of Wildlife Management*, vol. 63, no. 3, Jul. 1999.

[22] M. J. Gardner and D. G. Altman, "Confidence intervals rather than p values: estimation rather than hypothesis testing," *British Medical Journal (Clinical Research Edition)*, vol. 292, no. 6522, pp. 746–750, March 1986.

[23] J. Cohen, *Statistical power analysis for the behavioral sciences : Jacob Cohen.*, 2nd ed. Lawrence Erlbaum, Jan. 1988.
[24] ——, "The Earth is round ($p < 0.05$)," *American Psychologist*, vol. 49, no. 12, pp. 997–1003, 1994.
[25] M. Čadík, R. Herzog, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "New measurements reveal weaknesses of image quality metrics in evaluating graphics artifacts," in *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, vol. 31. ACM, 2012, pp. 1–10.
[26] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.

Ondřej Karlík Ondřej is currently a Ph.D. student at the Faculty of Mathematics and Physics, Charles University in Prague. He has been pursuing the Ph.D. degree since 2011. Before that he received the bachelor and master degrees in Computer graphics from Faculty of Electrical Engineering at Czech Technical University in Prague in 2009 and 2011, respectively. He is interested in both the theory of realistic image synthesis and its practical applications.

Martin Růžička Martin received the bachelor and Master degree in Computer Science from Charles University, Prague in 2009 and 2012. During his studies he was engaged in real-time computer graphics and afterwards he started to work for Take2K Czech as a game programmer.

Václav Gassenbauer Václav received the Master degree (Ing.) in computer science from Czech Technical University in Prague in 2006 and the Ph.D. degree in the same field from the University of Rennes in 2011. Currently, he is working as an external researcher developing driver assistance systems in an automotive industry. He has been solving problems from image processing and computer vision in his work.

Fabio Pellacini Fabio is an Associate Professor of Computer Science at Sapienza University of Rome and Dartmouth College. His research focuses on algorithms and systems for artist-friendly 3d content creation. Prior to joining academia, Fabio worked at Pixar Animation Studios on lighting algorithms, where he received credits on several movie productions. Fabio received his Laurea degree in Physics from the University of Parma (Italy), and his M.S. and Ph.D. in Computer Science from Cornell University (USA). Fabio received a National Science Foundation CAREER award, and an Alfred P. Sloan Research Fellowship.

Jaroslav Křivánek Jaroslav is an associate professor at Charles University in Prague. Prior to this appointment, he was a Marie Curie postdoctoral research fellow at the Cornell University Program of Computer Graphics, and a junior researcher and assistant professor at Czech Technical University in Prague. Jaroslav received his Ph.D. from IRISA/INRIA Rennes and the Czech Technical University (joint degree) in 2005. His primary research interest is realistic rendering and light transport simulation.