5-1-2012

Toward Evaluating Progressive Rendering Methods in Appearance Design Tasks

Jiawei Ou
Dartmouth College

Ondrej Karlik
Charles University in Prague

Jaroslav Krivanek
Charles University in Prague

Fabio Pellacini
Dartmouth College

Dartmouth Digital Commons Citation
Ou, Jiawei; Karlik, Ondrej; Krivanek, Jaroslav; and Pellacini, Fabio, "Toward Evaluating Progressive Rendering Methods in Appearance Design Tasks" (2012). Computer Science Technical Report TR2012-715. https://digitalcommons.dartmouth.edu/cs_tr/356

Dartmouth Computer Science Technical Report TR2012-715

Toward Evaluating Progressive Rendering Methods in Appearance Design Tasks

Jiawei Ou[1] Ondřej Karlík[2] Jaroslav Křivánek[2] Fabio Pellacini[1,3]
[1]Dartmouth College [2]Charles University in Prague [3]Sapienza University of Rome

Abstract

Progressive rendering is becoming a popular alternative to precomputation approaches for appearance design tasks. Images created by different progressive algorithms exhibit various kinds of visual artifacts at the early stages of computation. We present a user study that investigates the effects of these artifacts on user performance in appearance design tasks. Specifically, we ask both novice and expert subjects to perform lighting and material editing tasks with the following algorithms: random path tracing, quasi-random path tracing, progressive photon mapping, and virtual point light (VPL) rendering. Data collected from the experiments suggest that path tracing is strongly preferred to progressive photon mapping and VPL rendering by both experts and novices. There is no indication that quasi-random path tracing is systematically preferred to random path tracing or vice-versa; the same holds between progressive photon mapping and VPL rendering. Interestingly, we did not observe any significant difference in user workflow for the different algorithms. As can be expected, experts are faster and more accurate than novices, but surprisingly both groups have similar subjective preferences and workflow.

1 Introduction

Appearance design, i.e. the editing of lights and materials, is fundamental in the creation of computer-generated imagery, with a significant impact on the final image look. Kerr and Pellacini [2009; 2010] showed that in these tasks designers proceed mainly by trial-and-error, reporting on average 5 to 10 minutes for the manipulation of point lights and single BRDFs, among the simplest of design tasks. These tasks would take even longer in scenes with complex geometry and animation. More importantly, the timings reported assume that the renderer is fast enough to provide the user with immediate feedback. In practice, however, rendering realistic lighting and materials in complex environments, including global illumination effects, takes at least a few minutes.

Precomputation-based approaches have been introduced to speed-up rendering in the context of appearance design of complex environments [Pellacini et al. 2005; Hašan et al. 2006; Ben-Artzi et al. 2006]. These methods either support only one design task or do not allow artists to move geometry or camera. More importantly, they are based on approximations that cannot guarantee that exactly the same image will be generated by the final renderer, possibly misguiding the designer's effort.

For this reason, progressive rendering is becoming a popular alternative for providing fast feedback in appearance design tasks. A progressive renderer avoids pre-computation completely. Instead, it gradually improves the image quality until it converges to the final image, while providing the user with visual feedback during the entire course of computation. During design tasks, the renderer is restarted each time a scene changes, providing instantaneous feedback. At the early stages of computation, though, the image can contain various kinds of visual artifacts, such as high- or low-frequency noise or banding, that can interfere with the design task. Which of these artifacts are least objectionable in appearance design is an open question that we strive to answer.

In this paper, we present a user study that investigates the effects of the different artifacts produced by progressive renderers on user performance in appearance design tasks. Out of the large variety of progressive rendering algorithms, we chose the following four: purely random unstratified path tracing (showing high-frequency uncorrelated noise), quasi-random path tracing (showing high-frequency correlated noise), progressive photon mapping (showing low-frequency noise) and virtual point lights rendering (showing banding). We chose these methods, because (1) they span different types of visual artifacts, (2) they converge to the final rendered image, (3) they are used in practice, and (4) they have no initial latency, thereby supporting frequent user interactions.

In our study, fourteen expert subjects and twelve novice subjects perform simple tasks involving lighting and material design, receiving feedback from each of the aforementioned algorithms. In the *light matching task*, the subjects are asked to adjust a single parameter (either position or size) of one area light to match the given target image. In the *material matching task* the subjects adjust one of the color, glossiness, or roughness parameters of an object. In the lighting and material *open trials*, the subjects are asked to choose their preferred design out of eight predefined configurations of lighting or materials respectively. Subjects work with four scenes of varying complexity and extent shown in Fig. 1. We collect quantitative and qualitative data by recording all user interface actions and by asking subjects to fill out questionnaires, collecting ratings, rankings, and comments on each progressive renderer. By analyzing this data, we draw the following conclusions:

- Both path tracers are strongly preferred to progressive photon mapping and VPL rendering. As suggested by the time to completion and the algorithm rating/ranking, in appearance design tasks, users can cope better with the high-frequency artifacts of the path tracers, than the low-frequency noise and banding of progressive photon mapping and VPL rendering.
- The random and quasi-random path tracing are not systematically preferred to one another; the same holds between progressive photon mapping and VPL rendering.
- The four different algorithms do not result in any significant difference in user workflow.
- While experts are faster and more accurate than novices, the two groups have surprisingly similar preferences for the algorithms and exhibit similar workflow.

Although the results of the study apply strictly only to our scenes and tasks, we believe that they provide a valuable insight into developing progressive algorithms that specifically target appearance design.

2 Progressive Rendering Algorithms

We are interested in appearance design tasks in the context of realistic imagery. A wide range of rendering algorithms have been implemented to support realistic rendering (see [Pharr and Humphreys 2010] for a recent review), only a limited number of which were included in our study. The following criteria guided our selection.

- *Different types of image errors.* We include the algorithms that exhibit different types of image errors, including high-frequency noise, low-frequency splotches and banding.

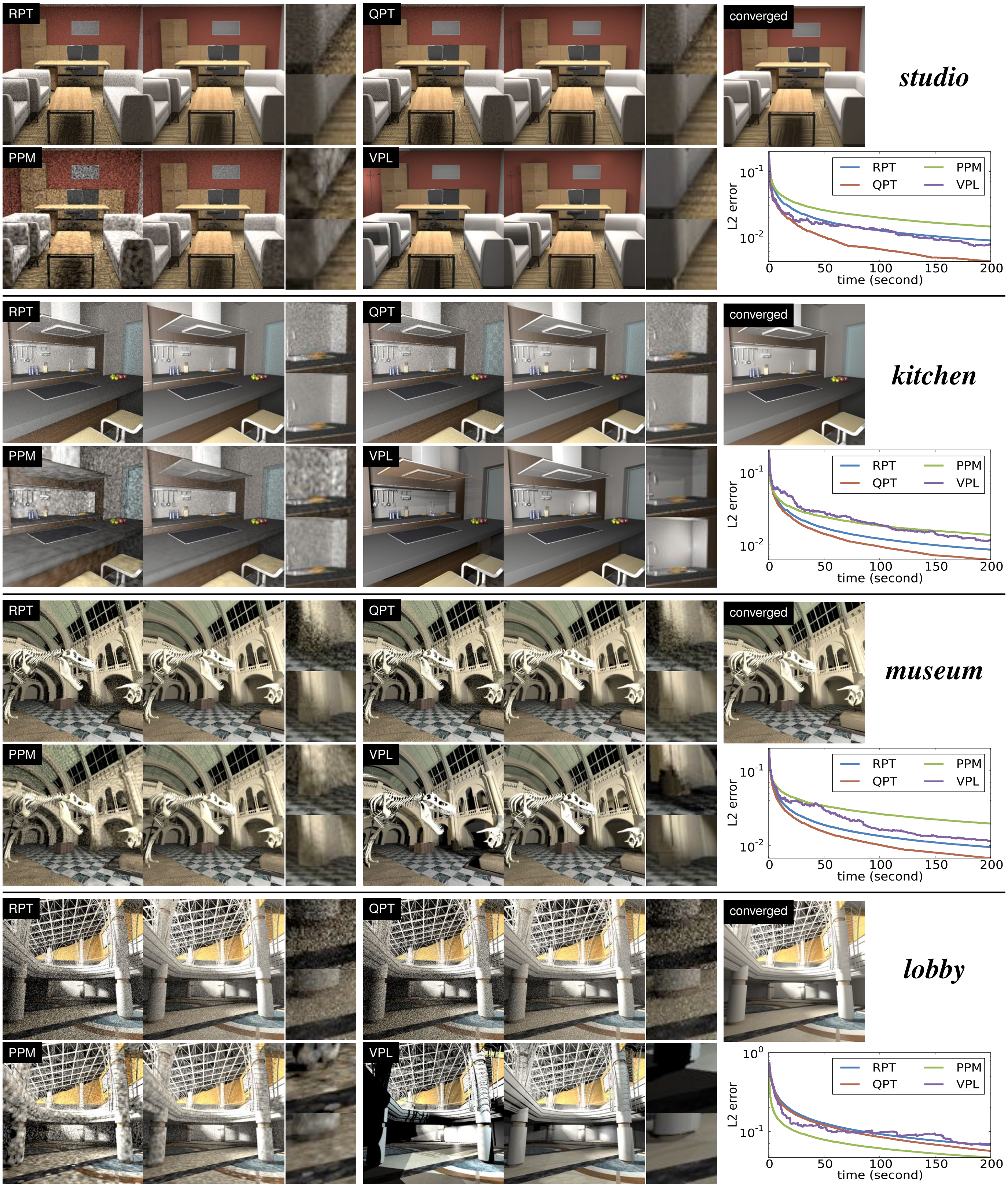


Figure 1: *We study four progressive rendering algorithms: random path tracing (RPT), quasi-random path tracing (QPT), progressive photon mapping (PPM), and virtual point light rendering (VPL). For each algorithm and each scene, we show the fully converged image and two partially converged images captured after 1 second and 10 seconds. Zoomed-in insets from the partially converged images show the algorithms' artifacts in detail. We also show a convergence graph of* L^2 *error plotted against time for the first 200 seconds.*

- *Convergence.* Recent real-time global illumination algorithms [Ritschel et al. 2009; Kaplanyan and Dachsbacher 2010] achieve high performance at the cost of approximations that may compromise image quality. To avoid affecting our results, we only consider algorithms that are known to converge to the correct solution, including both biased and unbiased techniques.
- *Used in practice.* We include algorithms that are widely used in industry and academia.
- *Minimum latency.* To support interactivity, rendering must restart immediately after any user action.

Based on these criteria we select the following four algorithms for our study: 1) random path tracing, 2) quasi-random path tracing, 3) progressive photon mapping, and 4) virtual point light (VPL) rendering. Fig. 1 shows the algorithms as they converge on the scenes used in our study. Note that while the different algorithms converge at similar speed according to the L^2 image error shown in the graphs, they exhibit very different artifacts. All images are rendered at 512×512 pixels and fully converged in a couple of minutes on a 27” iMac with a 4-core 2.93 GHz Intel Core i7 processor and 16 GB RAM. The supplemental video shows a screen capture of these algorithms as they converge on the chosen scenes. To ensure that our results can be reproduced, we include full source code for all the algorithms as supplemental material.

Path tracing Path tracing has recently gained popularity in industry practice for its ease of use and robustness [Křivánek et al. 2010a]. Random path tracing is a Monte Carlo solution of the rendering equation [Kajiya 1986] that provides unbiased results. A quasi-Monte Carlo version based on strictly deterministic number sequences can provide faster convergence in some situations [Keller 2003]. Non-converged images generated by path tracing exhibit unstructured high-frequency noise (pure random version) or structured high-frequency patterns (quasi-random version). Bidirectional path tracing [Veach and Guibas 1995] was not included in the study because it exhibits the same kind of error as a path tracer in our scenes.

Our implementation of the pure random path tracing algorithms follows Pharr and Humphreys [2010]. The quasi-random path tracer uses parametric quasi-Monte Carlo integration where one Sobol sequence is used for the entire image as in iray [Mental Images 2010]. By including both the random path tracer with no stratification and the quasi-random version which is stratified by nature, we are able to compare how stratification affects the performance of the two algorithms.

Progressive photon mapping Progressive photon mapping [Hachisuka et al. 2008] was introduced as a modification of the popular and widely used photon mapping method [Jensen 2001]. It is more robust than path tracing in some difficult lighting situations but has a slower asymptotic convergence rate [Knaus and Zwicker 2011]. At the initial stage, the algorithm generates smoother images than path tracing, which are affected by low-frequency noise. As the algorithm progresses, the noise becomes more fine-grained and decreases in magnitude.

Our implementation is based on stochastic progressive photon mapping [Hachisuka and Jensen 2009] where we connect photons to camera paths of one segment in length. In each iteration, we cast one camera ray per pixel, after which we trace 80,000 photons and connect them to the camera ray hit points. The photon mapping takes care of both direct and indirect illumination. The initial lookup radius was set manually; we used $\alpha = 0.7$ for the radius shrinking coefficient.

Practical implementations of *classic* photon mapping often compute direct illumination separately and only use the photon map for indirect illumination [Jensen 2001]. In our stochastic *progressive* photon mapping implementation, on the other hand, we use the photon map for both direct and indirect illumination evaluation for the following reasons. First, we wanted to conform to the original progressive photon mapping paper [Hachisuka et al. 2008]. Second and more importantly, we wanted to keep the four algorithms as clearly separated as possible by computing both direct and indirect lighting with the same method. This has the benefit that the artifacts specific to each method are displayed clearly and do not depend on the relative contribution of direct and indirect illumination in the particular view of the scene.

Virtual point lights VPL rendering [Keller 1997] is the basis of a number of real-time global illumination solutions. The algorithm is popular because smooth images free of any noise can be obtained quickly. This comes at the price of energy losses and banding artifacts, which can be disturbing especially at the early stages of progressive computation. Křivánek et al. [2010b] show that this method is not well suited to material design of highly glossy surfaces, but provides an acceptable approximation for low-gloss surfaces that we use in our study.

Our implementation of progressive VPL rendering adds illumination from one VPL in each iteration. We use the same number of virtual lights for direct and indirect illumination. Moderate clamping of light contributions is used to suppress artifacts due to weak singularities [Kollig and Keller 2004].

Implementation details All our algorithms are based on an optimized ray tracing engine and are fully parallelized. All renderers run asynchronously from the user interface to ensure that the latter is not blocked. To simplify the algorithms’ comparison, we deterministically stop all paths at length three. While this introduces bias in the solution, doing so suppresses the different impact that Russian roulette-based path termination may have on the convergence of the compared algorithms, and makes sure the amount of energy transfer computed by each algorithm in each render pass is identical. This ensures a fair comparison between all algorithms. Non-diffuse reflection is only computed at the first bounce from camera, while all other light bounces only consider the Lambertian component of the BRDF. This solution was chosen for compatibility with VPL rendering that does not easily support glossy reflection at the VPL [Hašan et al. 2009]. Moreover, excluding glossy reflections after the first bounce and limiting the path length are common approaches to improve the renderer efficiency in production practice.

Other methods We also tested a progressive version of Lightcuts [Walter et al. 2005], where the light cut at each pixel is refined progressively. This algorithm was not included in the study because it produces visually distracting artifacts and requires a large amount of memory, making it unsuitable for practical usage. Precomputation based approaches for interactive lighting or material design were also considered. These methods are based on moving some of the expense of rendering into a pre-computation stage [Pellacini et al. 2005; Hašan et al. 2006; Ben-Artzi et al. 2006; Sun et al. 2007; Ragan-Kelley et al. 2007; Ben-Artzi et al. 2008; Cheslack-Postava et al. 2008]. We did not include these algorithms in the study, because they do not support global illumination at all or cannot guarantee convergence to the correct image.

3 Experiment

Goal We seek to evaluate the *effectiveness of different progressive rendering algorithms for appearance design tasks*, namely light and material editing. Specifically, (1) we want to measure how efficiently users can perform editing tasks using different progressive rendering algorithms, and (2) we want to understand how the different artifacts and convergence behavior exhibited by each algorithm affect the way users perform appearance design tasks.

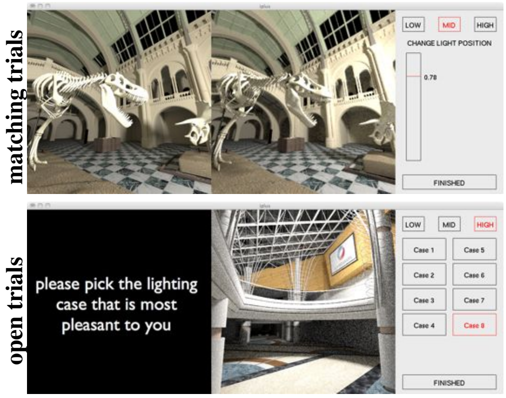
Subjects Fourteen expert subjects and twelve novice subjects participated in the experiment. All subjects had normal or corrected-to-normal vision. Subjects in the expert group perform most of their daily work using graphics design software. Most of them work in architectural or product visualization, using commercial 3D design software, such as 3ds Max and Maya, and have experience with multiple renderers, such as V-Ray and mental ray. All our experts subjects are capable of producing photorealistic images. Subjects in the novice group have limited experience with 3D design software. Most of them have never performed any appearance design before. However, the simplicity of the user interface and the design task we chose makes their performance in our experiments solely dependent on the behavior of the rendering algorithms, giving us clear measurements of algorithm qualities. We chose to include novices along with experts, because we believe they are more likely to need precise feedback from the rendering algorithms, and because our long-term goal is to make image synthesis more accessible to non-experts. Furthermore, comparing the performance and workflow of subjects of very different background gives us further insights on the generality of our results.

Scenes We designed our scenes to reflect the most common cases that professional lighters encounter. All scenes are lit by one area light and tone mapped with a fixed exposure-gamma algorithm ($\gamma = 2.2$). Materials are represented as a sum of a Lambertian diffuse lobe and a Blinn-Phong specular lobe. We positioned the camera to ensure that the light source was not visible in the rendered image. The *Studio* (104751 triangles) and *Kitchen* (183997 triangles) are relatively simple scenes lit mainly by direct and indirect illumination, respectively. The *Museum* (745944 triangles) is a more complex scene with strong direct and indirect lighting contribution. The *Lobby* (628478 triangles) is our most complex scene in terms of geometric detail, shadowing, and indirect effects. We chose diffuse and moderately glossy surfaces because these are the cases most commonly encountered in professional lighters' practice. In addition, including surfaces with more substantial gloss would handicap the VPL algorithm, which is unable to render them faithfully [Křivánek et al. 2010b].

Reducing complexity In this experiment, our focus is on measuring how the users are affected by the progressive algorithms themselves, rather than by the usability of the user interface. For this reason, we asked subjects to perform drastically simplified appearance design tasks.

Trials In *matching trials*, subjects were asked to match a given image by adjusting only one parameter corresponding to a property of the light or material. These simple tasks are representative of the basic editing operations that users perform with traditional user interfaces, where artists typically work by adjusting one parameter at a time, as shown in previous studies [Kerr and Pellacini 2009; Kerr and Pellacini 2010]. The matching trials allow us to quantitatively measure subjects' performance, while providing them with a clear goal. The supplemental video includes example user workflows for the matching tasks.

In *open trials*, we asked the subjects to make a subjective choice from a fixed set of predefined designs, each affecting many light or material properties. These operations are akin to picking the preferred option from a catalogue or by means of an image-based navigation interface. The open trials allow us to observe how subjects explore possible lighting/material configurations, which is a more natural task than matching. The user interfaces for the matching and open trials are shown in Fig. 3.



please pick the lighting case that is most pleasant to you

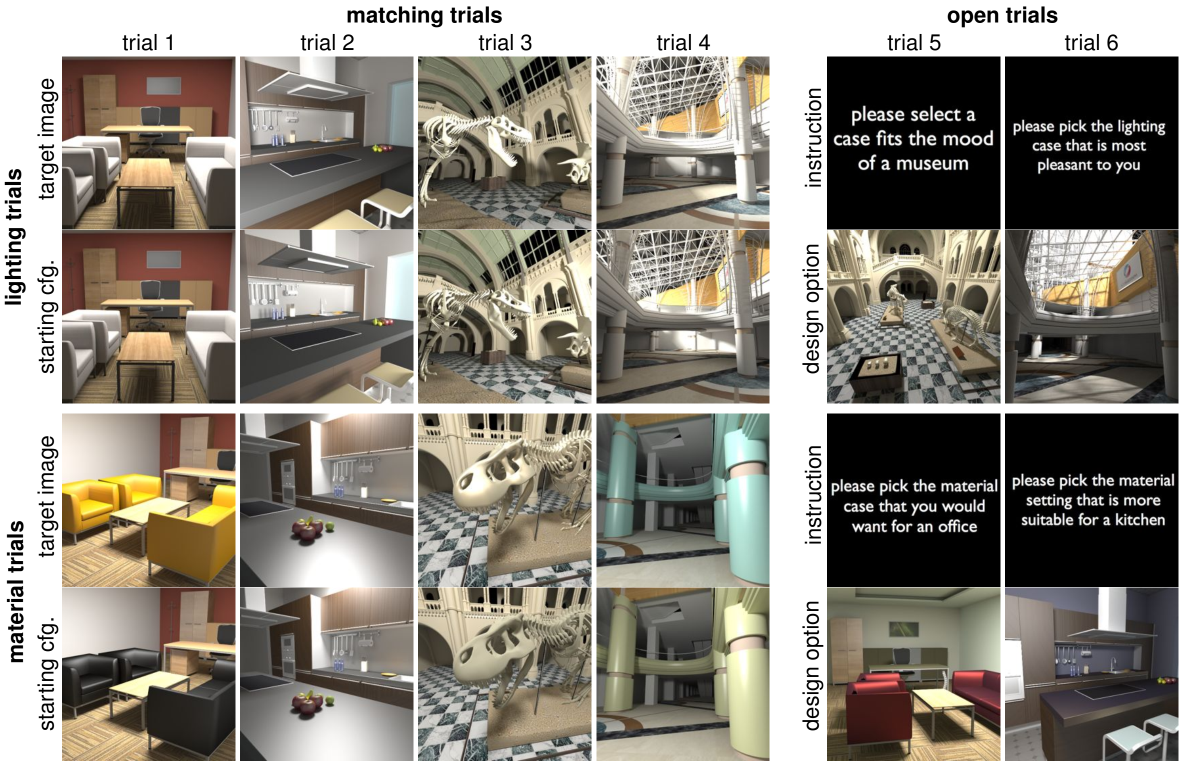
Figure 3: *User Interface for matching (top) and open (bottom) trials. Left: Target image or task instruction. Center: Progressive rendering. Right: Controls.*

Fig. 2 shows the starting and target configurations for the matching trials, as well as one of the predefined designs and the task description for the open trials. In the *light matching trials*, the subjects were asked to move or resize the light, while in the *material matching trials* they were asked to change the diffuse and specular intensity, specular roughness, or the diffuse hue. In each task, only one parameter can be changed by means of a slider. In the *open trials*, the subject were asked to choose from eight designs that we created by randomly changing either the light parameters (position, orientation, size, intensity) or the material parameters (diffuse and specular colors, and roughness). We chose a wide-angle camera view for the lighting tasks, while for material tasks we focused the camera on the edited object. To alleviate learning effects, we randomized the various starting and target configurations when the subject moves from one algorithm to another. To provide the subjects with a tradeoff between the image quality and speed, they can pick one of the three image resolutions: 128×128, 256×256, and 512×512 pixels.

Questionnaire After performing all trials with all algorithms, subjects are asked to complete a questionnaire where they rated each algorithm on a scale from 1 to 5 in the four categories corresponding to preference in lighting matching/material matching/open trials, and overall preference. Subjects also strictly ranked the algorithms in each of these categories. Immediately after finishing the trials using each algorithm, subjects were asked to leave free-form comments on their workflow and rate the subjective quality of the image they created. For subjects whose native language are not English, we translated the questionnaire in their language to allow them to faithfully express their opinion. For reproducibility, we include copies of the questionnaires as supplemental material.

Procedure The study consisted of four sessions, one for each algorithm (described in detail in Section 2). We randomized the order of algorithms given to each subject. In each session, the subject performed the following trials in order: 4 light matching trials, 2 light open trials, 4 material matching trials, and 2 material open trials. Before the study, we trained each subject individually to allow for questions and to accommodate each subject's learning needs. The instructor verified that the subject understood the task, and answered the subjects' questions. Once the study began, all user interface actions were recorded.

We conducted the study in two separate labs. The subjects performed all their trials in a controlled lighting environment with neg-



please select a case fits the mood of a museum

please pick the lighting case that is most pleasant to you

please pick the material case that you would want for an office

please pick the material setting that is more suitable for a kitchen

Figure 2: *Trials included in our study. Lighting matching trials: Subjects are asked to change the light size (trials 1) or adjust its position (trials 2 to 4) to match the target image. Material matching trials: Subjects are asked to change the brightness of the couch (trial 1), the roughness of the counter (trial 2), the highlight intensity on the dinosaur skull (trial 3), or the hue of the lobby pillar (trial 4). Lighting and material open trials: Subjects are asked to choose a pre-defined lighting or material design that fits a text description (trials 5 and 6).*

ligible ambient lighting, to simulate typical working conditions of artists and maximize visibility of the screen. The study was run with screens at resolution of 1600×900, at approximately 1 foot from the subject. All rendered images were 512×512 pixels on screen covering an area of roughly $4 \times 4 = 16$ square inches. We used an iMac with a 4-core 2.93 GHz Intel Core i7 processor and 16 GB RAM as our reference machine and synchronized the framerate of other machines to match its performance.

4 Analysis

We present our results in two parts. First, we analyze the output of the rendering system as the subjects proceed through each trial. Second, we compile the subjects' feedback from the questionnaires. Unless stated otherwise, tests for statistical significance are computed with repeated measures analysis of variance (ANOVA) [Stevens 2002]. The ANOVA for ranking data is conducted using the *Kruskal–Wallis* test, which is a nonparametric alternative that does not rely on an assumption of normality. A p value below 0.1 indicates a 90% confidence that the two population means differ given the measure of the sample. In all figures, error bars represent standard error. All L^2 errors are computed in gamma corrected RGB space.

Matching error To evaluate user performance during the matching trials, we compute the L^2 error between the rendered image and the target image (see Fig. 4). Given the progressive nature of the algorithms, the rendered image keeps refining if there is no user interaction. At each user action, the renderer restarts and the error jumps up, which explains the fluctuations in the individual graphs in Fig. 4, left. Note that the error shown here is not a measure of how well the target would be matched after convergence; it merely shows the difference between the images at a given point in time.

To illustrate how the algorithms support user workflow, we average all user errors corresponding to the same trial in Fig. 4. While this cannot be used for a quantitative analysis, it helps to highlight the general subject behavior. Most of the subjects, including novices and experts, were able to closely match the target image in all but one matching trial. In this material trial 2, subjects were asked to match the material roughness on a flat counter. This task was found to be too hard for novice subjects with all rendering algorithms. However, expert subjects did not have such problems.

Time to completion Given enough time, all the algorithms eventually converge. The true difference between them is in the early stages of computation, so time to completion is an important measure of an algorithm's performance. In particular, the algorithm that provides the most useful feedback for appearance design should show shorter time to completion.

Fig. 5 shows the average time to completion for each type of trial for each subject group. In light matching trials, all subjects were able to finish sooner using one of the path tracers than using the virtual point light algorithm ($p < 0.04$). In addition, expert subjects also finished tasks sooner using the progressive photon mapping than using the virtual point light algorithms ($p < 0.06$), but such a trend is not shown in the novice group. This indicates that regardless of subjects' experience, path tracers can provide more useful feedback, while progressive photon mapping is useful but limited to the experts. In general, subjects try to match shadows/highlight first, and only if these were not sufficiently salient, they consider other lighting features. In questionnaires, subjects left comments such as "for lighting look at the shadow, for material look at the color and brightness", and "if there is shadow in the figure, it is much easier than without shadow".

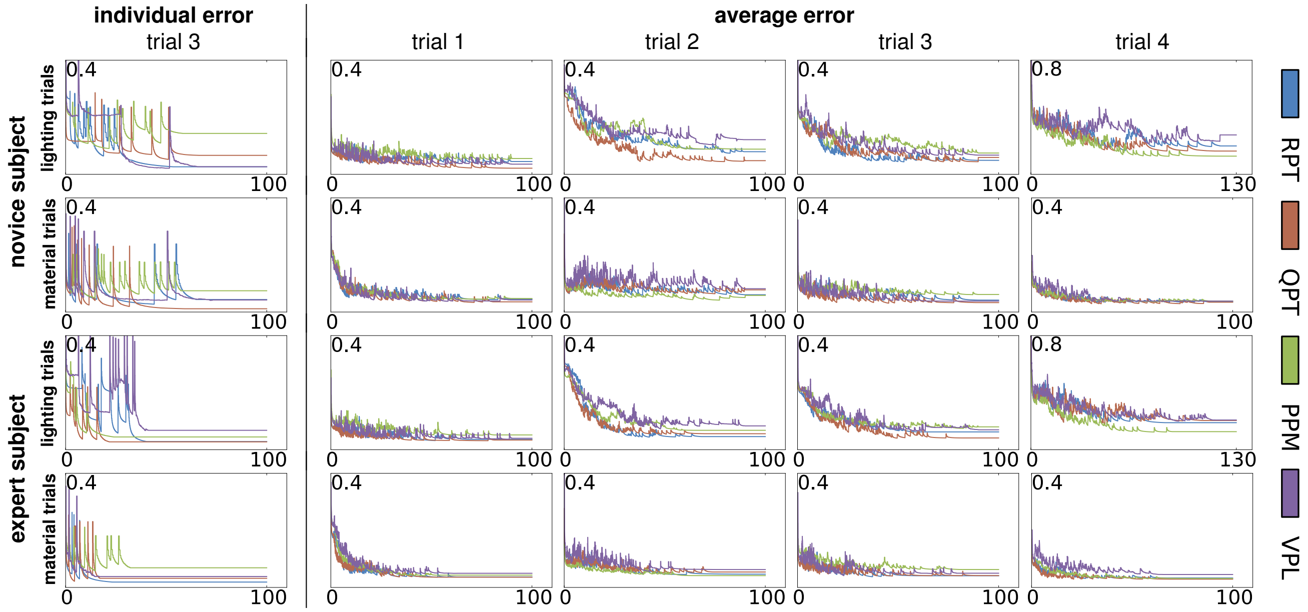


Figure 4: *L^2 error (vertical axis) over time (horizontal axis, in seconds) for matching trials. Left: Error for an example subject. Right: Average error across subjects. A complete set of L^2 error plots for each individual subject is provided in the supplemental material.*

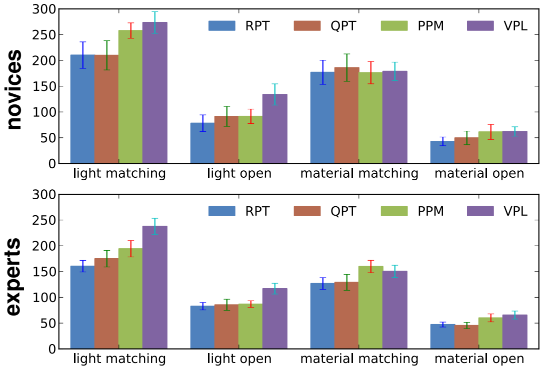


Figure 5: *Average time to completion.*

For material matching trials, no clear trend is present. All subjects finished in roughly the same time. We attribute this to the fact that in light matching trials subjects needed an overview of the lighting of the entire scene to make a decision. In contrast, in material editing tasks, subjects had a limited view to the scene and a limited feedback sufficed to perform the task, making the difference between algorithms less important. Comments in the questionnaires confirm this: "Tuning material is easy, the rendering of lighting is sort of slow".

In the open trials, subjects have lower average completion time than in the matching trials with all algorithms ($p < 0.01$). However, the meaning of time to completion is less well-defined since it depends on the standard of judgment the subject has chosen. While novice subjects show no significant trend between algorithms, expert users spent much more time in both light and material open trials with VPLs than with the other three algorithms ($p < 0.05$).

In general experts subjects finished sooner than novice subjects when using the random path tracer in light matching trials ($p < 0.08$) and material matching trials ($p < 0.06$). A similar trend is also observed for quasi random path tracing, but with less certainty.

Scene complexity We found that in lighting design, scene complexity has a large impact on the user performance. The main reason is that complicated light paths affect the different algorithms in a different manner. For example, the lobby scene in lighting trial 4 has a higher matching error for all algorithms (see Fig. 4). In this case, progressive photon mapping performs much better but still shows artifacts that hinder users' ability to perform design tasks. However, in material design tasks, we found that the performance is independent of the geometric complexity, since subjects focus mostly on a small part of the environment.

Subjective image quality At the end of each trial, we asked the subjects to rate their work on the scale from 1 to 5, where 1 means the worst and 5 means the best. Matching trials were rated in terms of how closely the subjects are able to match the reference image. Open trials were rated in terms of how satisfied they are with their choice. Fig. 6 shows average rating for each kind of trial and for each user group. In the light matching trials, all subjects on average rated their work better when using the two path tracers ($p < 0.08$ for novices and $p < 0.01$ for experts[1]) suggesting that they perceived themselves doing a better job with these algorithms. However, for other three types of trials, no obvious trend is shown in the novice subject group. We attribute this to the fact that, unlike light matching, material matching and open trials are too subtle for novices to properly rate their work.

We also observe that expert subjects have substantially lower rating for tasks done with the virtual point light algorithm in light open trials compared to the other three algorithms ($p < 0.07$), meaning that expert subjects are unsatisfied with the feedback provided by the VPL algorithm when they need to get a general sense of the entire scene instead of matching specific lighting features. This is also confirmed in the questionnaire "Overall, it [VPL] is unpleasant to look at. I don't see any advantage." "It [VPL] made me wait longer. Previews kept changing."

Workflow in matching trials In matching trials, most subjects employed a simple search and refine approach. Subjects would first click different random positions on the slider. Once they found a rough value, they began to perform smaller adjustments to find the

---

[1] This is the upper bound of the p value for the ANOVA of (RPT vs. PPM), (RPT vs. VPL), (QPT vs. PPM) and (QPT vs. VPL).

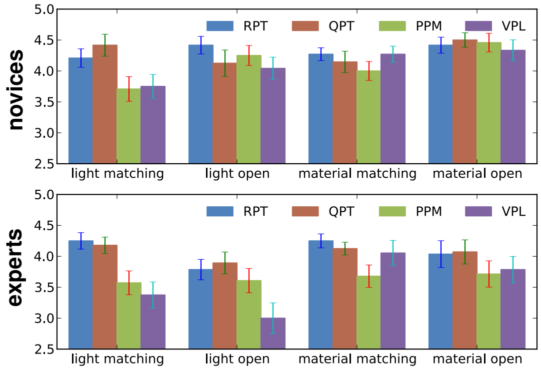


Figure 6: *Subjective Image Rating. Experts are less satisfied with their results, indicating that they are more precise in appreciating appearance differences.*

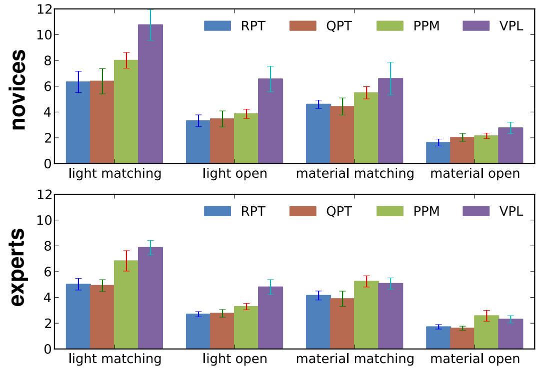


Figure 7: *Average time between user interactions in seconds. For matching tasks, it is the average interval between two clicks on the slider. For open tasks, it is the average interval between changing the designs. Experts take slightly less time than novices.*

target. If they received fast feedback, they tended to click more. Some of the users would repeatedly click the slider bar to simulate a dragging effect. This is confirmed by looking at the individual videos for each subject. Moreover, the behavior of novice subjects and expert subjects do not differ. Fig. 7 shows that subjects have shorter interaction interval when using the two path tracers compared to progressive photon mapping and VPL ($p < 0.05^1$), indicating that path tracers provide useful feedback at the early stage, which allows subjects to make decisions faster. By observing the user interactions, we found that the workflow is similar in the two groups.

Workflow in open trials In the open trials, subjects explored the pre-defined design options. At first, they browsed through all the options and waited a short time to get a sense of how each design looks like. After this first round, users would go back to a few of the designs that interested them and made a final decision by going back and forth amongst them. Less time was spent in each configuration than in the matching trials ($p < 0.02$) (see Fig. 7). This indicates that when the task is more subjective and only requires a high-level decision, progressive algorithms are able to provide a very quick preview that users find helpful in this context. Trends similar to matching trials are observed, two path tracers have shorter interaction time in both light open trials ($p < 0.09$) and material open trials ($p < 0.07$). Both groups performed their tasks using a similar workflow.

Resolution switching Subjects were allowed to change image resolution during each trial. Most of them started at the lowest resolution and later switched to higher resolution for finer tweaking. If they felt the result needed more than a small adjustment to match the target, they switched back to a lower resolution. One comment says: "I have been using low resolution to get an approximation and then high resolution for more precision". Another says: "The more I approached the target, the more important the render quality was so that I could compare the image details."

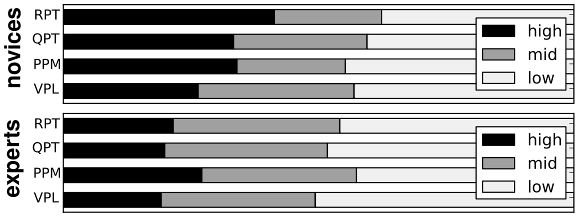


Figure 8: *Proportion of time users spent at each resolution. Experts spend less time at high resolution.*

We notice that the subjects spent more time working in low and medium than in high resolution (see Fig. 8). Most subjects only switched to high resolution at the end to validate their work; a few subjects never used the high resolution mode in matching tasks. This trend is even more obvious in the open trials: "[In open trials,] I switched to low resolution and was looking mainly at the colors". We conclude that most users do not need a fully converged, high-quality image to make their choice and having a faster low-resolution mode is useful for all progressive algorithms. The supplemental material shows graphs of the time spent at each resolution for individual subjects.

In general, compared to novices, experts were more likely to stay in the low and middle resolutions, suggesting that they were comfortable making more decisions based on lower quality imagery, probably due to their familiarity. Another interesting observation is that expert subjects stayed in high resolution longer when using photon mapping compared to other three algorithms. A few users even stated in the questionnaire that they need to use higher resolution with progressive photon mapping. "I had to increase the quality.", "I nearly did not use the low resolution.", "I immediately switched to MID or HIGH (when using PPM)."

Algorithm ratings and rankings Subjects rated each algorithm in all four types of trials. Fig. 10 summarizes average ratings. In the overall rating, novice subjects rated the two path tracers higher than the other two methods ($p < 0.05$). However, when observing the individual trials, this preference is not significant, except in the light matching trials ($p < 0.01$).

On the other hand, the expert subjects show a clear preference toward the two path tracers in the ratings statistics for all trial types

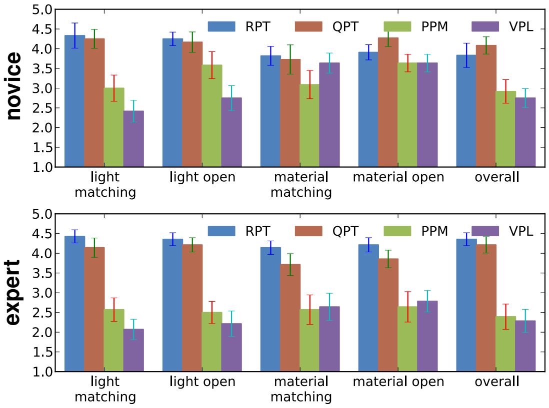


Figure 10: *Algorithm ratings. Experts have a stronger preference toward path tracing algorithms.*

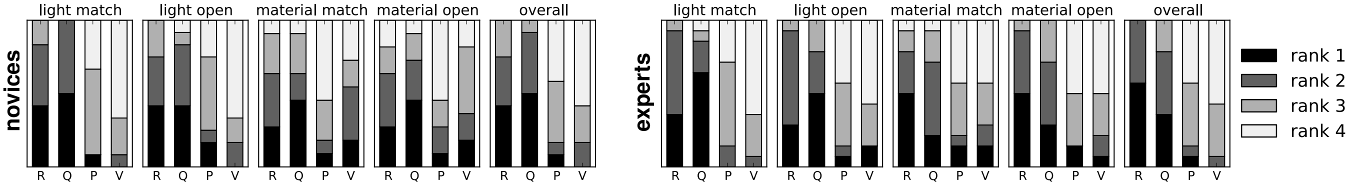


Figure 9: *Algorithm rankings. (R = RPT, Q = QPT, P = PPM, V=VPL)*

($p < 0.03$). This was mirrored by the comments in the questionnaires. For example, regarding progressive photon mapping, some subjects wrote: "it was hard to get an accurate match because of the artifacts even at high resolution", and "photon mapping generates too large splotches and it is hard to find the key areas for comparing the images (edge, shadow, highlight)", and finally "[I] have to wait to see small details in both light and material". Regarding VPL method they comment: "things change very slowly and may look very different before it converges", "it was easy to settle on a rough approximation, but I was never confident of the stability of the final choice", "I was annoyed by the blinking from VPLs".

There is no systematic difference between the two path tracers in terms of ratings. Subjects' comments confirm this, for example "random path tracing and quasi-random path tracing were quite similar except quasi-random was slightly clearer", "RPT+QPT - I did not notice any difference. I found them relatively fast and accurate", and "Differences between QPT a RPT are not noticeable. They can provide an overall preview very swiftly.".

Subjects were also asked to rank each algorithm in all four types of trials. While rating can have ties, rankings are a forced choice. Fig. 9 shows the stacked frequencies of rankings. In the novice group, the two path tracers were ranked higher than the other two methods in the overall ranking and in the two lighting trials. In the material trials, progressive photon mapping received the lowest ranking among all algorithms ($p < 0.06$), but no systematic difference is shown between the two path tracers and VPL method.

On the other hand, in the expert group, subjects consistently ranked the two path tracers higher than the other two algorithms in overall ranking statistics ($p < 0.03$), but there is no systematic difference between the two path tracers in terms of ranking. The strong preference for path tracing is one of the main results of our work.

5 Discussion and Future Work

In this section, we summarize the major findings of our study. Before continuing, we want to remind the reader that strictly speaking our observations only apply within the boundary of the tested cases, just like all user studies. At the same time, given that the observed trends were consistent for the four different scenes included in our study, we believe that they are general enough to apply to many other scenes as well.

Progressive renderers Our most prominent result is the poor performance of progressive photon mapping and virtual point light rendering compared to the two path tracers. We found that while subjects are able to perform simple appearance design tasks well using all algorithms, their performance is better with path tracers. This trend can be observed objectively in the time to completion. Moreover, the path tracers received higher rating and ranking, indicating that the users actually prefer working on appearance design using these methods.

In our tests, the random and quasi-random path tracers had similar performance and subjects did not consistently prefer one over the other. This result is surprising especially because our random path tracer did not use any stratification while the quasi-random path tracer is stratified by nature. Similarly, we did not observe any systematic preference of progressive photon mapping over virtual point light rendering or vice versa. This suggests that the high-frequency errors exhibited by the path tracers are easier for subjects to cope with than the low-frequency errors or banding shown by the other methods. This trend is more obvious in the expert group than in the novice one.

Experts vs. Novices In general, expert subjects are more efficient than novices. This is confirmed by the statistics of average time between interactions and time to completion, and also by watching the captured videos. Expert subjects appreciate appearance differences better than novice subjects. The differences between the algorithms' artifacts substantially influence expert subjects' decision in the final rating and ranking. Moreover, the statistics of the expert group usually have a clearer trend with lower variance, meaning that in general, expert subjects behave more consistently. The statistics of the novice group tend to have higher variance, showing that novices are less predictable. Unlike the expert subjects, the algorithm ratings of novice subjects do not reflect a clear preference. But when it comes to ranking which subjects are forced to choose, novice subjects made decisions similar to expert subjects.

Common workflow Our subjects exhibit common workflow patterns. In matching tasks, they generally employ a search and refine approach, first finding a rough position around the target and then refining it by making small changes. This workflow is independent of the progressive renderer used as well as the subject experience level. A similar trend was found for open trials, but with quicker user decisions. Furthermore, subjects are willing to sacrifice image resolution for faster feedback in the initial search, while they switch to higher resolution as they refine.

Initial user feedback The search-and-refine workflow together with the resolution switching behavior suggest that in appearance editing subjects favor algorithms that provide immediate feedback on the overall scene look while refining the details later.

Limitations As in all user studies, the main limitation of our work is the scope of our investigation, in terms of the algorithms and of the lighting and material editing tasks we have explored. Moreover, we have only explored a fraction of the possible lighting models, material models, and scene settings. In material trials, we chose what we believe are the most common material design tasks, but we acknowledge that the results may not hold for very different tasks such as spatially varying material design or texture selection.

Future work These limitations suggest clear directions in expanding the scope of our work in the future. At the same time, we feel that the observed trends are general and likely to be confirmed by further studies, especially considering that similar trends are observed in different subjects' groups. We believe that a more fruitful avenue for further exploration is the development and testing of appearance design user interfaces that work in conjunction with progressive renderers, rather than the current interfaces that fundamentally assume the renderer has perfect image quality. Kerr and Pellacini [2009; 2010] showed that the choice of user interface has a significant impact on user performance when coupled with a real-time renderer. The question we are interested in is how to design effective interfaces that can help users in design tasks when the feedback is given by, for example, a progressive path tracer.

6 Conclusion

This paper presents a first step toward the evaluation of progressive rendering algorithms in the context of appearance design. By performing a series of matching and open trials and by collecting subject evaluation in questionnaires, we have measured how different progressive rendering algorithms aid both novice and expert subjects in performing specific lighting and material design tasks. In comparing path tracing with progressive photon mapping and virtual point light (VPL) rendering, we found the former to perform better in terms of objective and subjective measures. This trend was common in both subject groups, further strengthening the results. The main differences between subject groups were that experts were faster and more precise overall.

While, as in any user study, we acknowledge that our measurements are only strictly valid within the context of our experiment, we believe that the main trends found in our study generalize to other scenes and appearance design tasks. In addition, we expect that our experiment design will be used as the basis for further explorations of the effectiveness of progressive rendering algorithms. In the future, we are interested in extending our study to include more sophisticated appearance tasks and different user interfaces. More importantly though, we are interested in investigating how to design user interfaces that work in conjunction with progressive renderers.

References

BEN-ARTZI, A., OVERBECK, R., AND RAMAMOORTHI, R. 2006. Real-time brdf editing in complex lighting. *ACM Transactions on Graphics 25*, 3 (July), 945–954.

BEN-ARTZI, A., EGAN, K., RAMAMOORTHI, R., AND DURAND, F. 2008. A precomputed polynomial representation for interactive brdf editing with global illumination. *ACM Transactions on Graphics 27*, 2 (Apr.), 13:1–13:13.

CHESLACK-POSTAVA, E., WANG, R., AKERLUND, O., AND PELLACINI, F. 2008. Fast, realistic lighting and material design using nonlinear cut approximation. *ACM Transactions on Graphics 27*, 5 (Dec.), 128:1–128:10.

HACHISUKA, T., AND JENSEN, H. W. 2009. Stochastic progressive photon mapping. *ACM Transactions on Graphics 28*, 5 (Dec.), 141:1–141:8.

HACHISUKA, T., OGAKI, S., AND JENSEN, H. W. 2008. Progressive photon mapping. *ACM Transactions on Graphics 27*, 5 (Dec.), 130:1–130:8.

HAŠAN, M., PELLACINI, F., AND BALA, K. 2006. Direct-to-indirect transfer for cinematic relighting. *ACM Transactions on Graphics 25*, 3 (July), 1089–1097.

HAŠAN, M., KŘIVÁNEK, J., WALTER, B., AND BALA, K. 2009. Virtual spherical lights for many-light rendering of glossy scenes. *ACM Transactions on Graphics 28*, 5 (Dec.), 143:1–143:6.

JENSEN, H. W. 2001. *Realistic Image Synthesis Using Photon Mapping*. AK Peters.

KAJIYA, J. T. 1986. The rendering equation. In *Computer Graphics (Proceedings of SIGGRAPH 86)*, 143–150.

KAPLANYAN, A., AND DACHSBACHER, C. 2010. Cascaded light propagation volumes for real-time indirect illumination. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*, ACM, I3D '10, 99–107.

KELLER, A. 1997. Instant radiosity. In *Proceedings of SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 49–56.

KELLER, A. 2003. Strictly deterministic sampling methods in computer graphics. In *ACM SIGGRAPH 2003 Courses, Course No. 44: Monte Carlo Ray Tracing*.

KERR, W. B., AND PELLACINI, F. 2009. Toward evaluating lighting design interface paradigms for novice users. *ACM Transactions on Graphics 28*, 3 (July), 26:1–26:9.

KERR, W. B., AND PELLACINI, F. 2010. Toward evaluating material design interface paradigms for novice users. *ACM Transactions on Graphics 29*, 4 (July), 35:1–35:10.

KNAUS, C., AND ZWICKER, M. 2011. Progressive photon mapping: A probabilistic approach. *ACM Transactions on Graphics 30*, 3 (May), 25:1–25:13.

KOLLIG, T., AND KELLER, A. 2004. Illumination in the presence of weak singularities. In *Monte Carlo And Quasi-Monte Carlo Methods*, 245–257.

KŘIVÁNEK, J., FAJARDO, M., CHRISTENSEN, P. H., TABELLION, E., BUNNELL, M., LARSSON, D., AND KAPLANYAN, A. 2010. Global illumination across industries. In *ACM SIGGRAPH Courses*.

KŘIVÁNEK, J., FERWERDA, J. A., AND BALA, K. 2010. Effects of global illumination approximations on material appearance. *ACM Transactions on Graphics 29*, 4 (July), 112:1–112:10.

MENTAL IMAGES, 2010. iray. http://www.irayrender.com/.

PELLACINI, F., VIDIMČE, K., LEFOHN, A., MOHR, A., LEONE, M., AND WARREN, J. 2005. Lpics: a hybrid hardware-accelerated relighting engine for computer cinematography. *ACM Transactions on Graphics 24*, 3 (Aug.), 464–470.

PHARR, M., AND HUMPHREYS, G. 2010. *Physically Based Rendering, From Theory To Implementation*, 2nd ed. Morgan Kaufmann.

RAGAN-KELLEY, J., KILPATRICK, C., SMITH, B. W., EPPS, D., GREEN, P., HERY, C., AND DURAND, F. 2007. The Lightspeed automatic interactive lighting preview system. *ACM Transactions on Graphics 26*, 3 (July), 25:1–25:11.

RITSCHEL, T., ENGELHARDT, T., GROSCH, T., SEIDEL, H.-P., KAUTZ, J., AND DACHSBACHER, C. 2009. Micro-rendering for scalable, parallel final gathering. *ACM Transactions on Graphics 28*, 5 (Dec.), 132:1–132:8.

STEVENS, J. 2002. *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum.

SUN, X., ZHOU, K., CHEN, Y., LIN, S., SHI, J., AND GUO, B. 2007. Interactive relighting with dynamic brdfs. *ACM Transactions on Graphics 26*, 3 (July), 27:1–27:10.

VEACH, E., AND GUIBAS, L. J. 1995. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of SIGGRAPH 95*, Computer Graphics Proceedings, Annual Conference Series, 419–428.

WALTER, B., FERNANDEZ, S., ARBREE, A., BALA, K., DONIKIAN, M., AND GREENBERG, D. P. 2005. Lightcuts: a scalable approach to illumination. *ACM Transactions on Graphics 24*, 3 (Aug.), 1098–1107.