The Perception of Lighting Inconsistencies in Composite Outdoor Scenes

MINGHUI TAN, Yale University
JEAN-FRANÇOIS LALONDE, Laval University
LAVANYA SHARAN, Massachusetts Institute of Technology
HOLLY RUSHMEIER, Yale University
CAROL O'SULLIVAN, Disney Research Los Angeles

It is known that humans can be insensitive to large changes in illumination. For example, if an object of interest is extracted from one digital photograph and inserted into another, we do not always notice the differences in illumination between the object and its new background. This inability to spot illumination inconsistencies is often the key to success in digital "doctoring" operations. We present a set of experiments in which we explore the perception of illumination in outdoor scenes. Our results can be used to predict when and why inconsistencies go unnoticed. Applications of the knowledge gained from our studies include smarter digital "cut-and-paste" and digital "fake" detection tools, and image-based composite scene backgrounds for layout and previsualization.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Display algorithms*

General Terms: Human Factors, Experimentation

Additional Key Words and Phrases: Lighting, Perception, Images

ACM Reference Format:

Minghui Tan, Jean-François Lalonde, Lavanya Sharan, Holly Rushmeier, and Carol O'Sullivan. 2015. The perception of lighting inconsistencies in composite outdoor scenes. ACM Trans. Appl. Percept. 12, 4, Article 18 (September 2015), 18 pages.
DOI: http://dx.doi.org/10.1145/2810038

1. INTRODUCTION

Combining different structures extracted from existing photographs to generate a novel composite scene has many practical applications. Photographs of existing physical scenes are often used as inspiration for applications such as designing architectural structures or creating sets for film production, for example. These photos are then combined together in various ways, thus allowing the designer or film director to evaluate alternative designs and to rapidly "previsualize" different concepts.

This work was supported in part by grant #1302267 from the US National Science Foundation.
The contact author can be reached at carol.osullivan@tcd.ie.
Authors' addresses: M. Tan and H. Rushmeier, Computer Graphics Group, Yale University, New Haven, CT 06520, U.S.A.; emails: tanminghui.ivy@gmail.com, holly.rushmeier@yale.edu; J.-F. Lalonde, Computer Vision and Systems Lab., Laval University, Quebec City, QC, G1V 0A6, Canada; email: jflalonde@gel.ulaval.ca; L. Sharan, Dept. of Brain & Cognitive Sciences and CSAIL, MIT, Cambridge, MA 02139; email: lavanya@csail.mit.edu; C. O. Sullivan, Disney Research Los Angeles, 1401 Flower Street, Glendale, CA 91201, U.S.A.; email: carol.osullivan@tcd.ie.

DOI: http://dx.doi.org/10.1145/2810038

To be a useful system, we require that the previsualization look natural and that any newly inserted structure does not appear out of place. Several factors might make the structure appear to "pop out": (1) bad segmentations (e.g., the structure was improperly segmented from its original image); (2) the scene semantics are wrong (e.g., the structure appears to be floating in midair); (3) mismatched geometry (e.g., it is not oriented properly with respect to existing structures); and/or (4) mismatched lighting conditions (e.g., shadows point in different directions). All these considerations must be taken into account by the artist when generating such previsualization scenes, which is an arduous task.

While powerful image editing tools already exist to account for problems 1–3 listed earlier (segmentations, semantics, and, to some extent, geometry), the same cannot be said of mismatched illumination conditions, which are very hard to modify after the fact as they require an accurate estimate of the shape and material properties of the object to transfer. Luckily, it has been shown that humans are notoriously bad at reasoning about the physical accuracy of lighting effects [Cavanagh 2005], so maybe illumination does not have to be matched exactly to generate a realistic composite.

Is there a point at which we do not notice illumination inconsistencies? Several studies have attempted to answer this question, but so far they have been performed in the lab with abstract or limited indoor stimuli, so it is not clear whether their results will generalize to the much more complex setting of real-world, outdoor scenarios. In this work, we focus on determining whether a resulting image composite will appear realistic, by analyzing the perception of directional lighting inconsistencies in composite outdoor scenes. To this end, we introduce two novel experiments. The first experiment employs composite images generated from real webcam images, where the lighting direction is known. The observations we made in this first experiment, and the limitations we encountered in using the webcam data, informed the design of a second experiment. This second experiment, from which we draw most of our final results, is based on a set of realistic synthetic outdoor images. In both cases, one structure in the scene exhibits varying degrees of illumination inconsistencies.

2. BACKGROUND

Inserting new objects into photographs of existing scenes is a common task in computer graphics. Successfully inserting new objects requires estimating the lighting in the scene and then rendering the synthetic objects according the estimated illumination [Karsch et al. 2011, 2014]. Many algorithms may be used to estimate the illumination in an image [Lalonde et al. 2012; Chen et al. 2011; Lopez-Moreno et al. 2013]. Algorithms have also been proposed to determine the realism of image composites [Reinhard et al. 2001; Lalonde and Efros 2007; Xue et al. 2012]. In particular, Xue et al. systematically examined the perceived realism of a composite by independently manipulating luminance, color temperature and saturation in a foreground image. Based on their findings, they developed algorithms for altering these quantities to create realistic composites. However, such techniques cannot alter the spatial patterns of light and dark regions produced by the original illumination in a scene.

Previously, several experiments have been conducted to explore human sensitivity to illumination changes. Koenderink et al. have studied how humans estimate light direction for a synthetic Gaussian landscape viewed from above [Koenderink et al. 2004], and for captured imagery of textured surfaces in the CURET database, viewed along the normal to the sample surface [Koenderink et al. 2003]. Observers could accurately estimate the illumination direction for the artificial surfaces but not for the images of real-world surfaces. Unlike in our work, Koenderink et al. do not examine the detection of inconsistencies—rather, they focus on the ability to estimate the lighting direction. Also, they only consider the case of viewing a surface or landscape normal to the surface. In our case, we consider a variety of surface orientations that are typically observed in outdoor scenes.

Ostrovsky et al. [2005] have examined the ability to detect lighting inconsistencies in simple synthetic test data (e.g., free-floating cubes). They argue that substantial anomalies can be introduced

into complicated scenes without detection, and present a series of "doctored" real-world photographs confirming this observation. They note that there is no evolutionary advantage to detecting inconsistencies in a scene—we act on local illumination information, so why would we have developed the ability to pick out these anomalies? They hypothesize that when compositing artifacts are visible in feature film, it may be because of the conflicting brightness levels, rather than the conflicting illumination directions. Furthermore, while they were able to illustrate the possibility of constructing scenes in which lighting inconsistencies are not noticeable, they did not provide rules for doing so.

Lopez-Moreno et al. [2010] studied the ability to detect lighting inconsistencies with the aim of quantifying the point at which the inconsistencies become noticeable. In their study, variations in lighting direction were limited to variations in a plane normal to the "vertical" direction in the scene. Furthermore, they only considered blob-like objects and physical scenes in which objects were organized to eliminate cast shadows, thus achieving a limited depth of field. Their study of isolated diffuse and specular objects indicated a threshold of a 30° (for light coming from the front) and 20° (for light coming from the back). They also found that detection was more difficult for textured objects, up to a threshold of 40°. The authors remark that this increased threshold was unexpected, since it would seem that inconsistencies would be easier to spot in natural textures. For our applications, we want to see whether their results hold for architectural-scale outdoor scenes. We also wish to see if the same thresholds apply to variations of light direction, with respect to the ground plane. Note that the set of light directions tested in their study naturally correspond to sun positions in the sky.

We extend the previous experiments in lighting inconsistency perception to the analysis of outdoor scenes that are of interest in applications such as film set creation and architectural design. In this context, we seek specific parameters that can be used in rendering and compositing. We are interested in the following questions:

(1) Can we quantify the noticeable light source inconsistencies for sun direction relative to view direction in outdoor scenes?
(2) Do we see the same thresholds for angle as observed by Lopez-Moreno et al. [2010] for the real and textured scenarios? Do we see the same backlit/frontlit effects?
(3) Do we see the same thresholds for angles vary with respect to height above the viewer (elevation) as for angular variations from left to right (azimuth)?
(4) Is the sensitivity for large structures in large depth-of-field images the same as for sets of synthetic or low depth-of-field objects?
(5) Is there an effect of whether light comes from the left or right of the camera?

3. EXPERIMENT I: REAL SCENES

We first explore the effects of illumination inconsistencies in composite scenes by using real imagery taken from webcam sequences. We chose to experiment with real images, since our motivating previsualization application involves compositing into natural images. In this section, we describe how stimuli were generated, the experiment performed on this data, and the results, and we discuss the insights we obtained and the limitations we encountered.

We selected 15 different scenes from the Webcam Clipart Database [Lalonde et al. 2009]. This database was chosen due to the availability of the sun position with respect to the camera at each frame of a sequence. Figure 3 shows five example scenes used in this experiment. Although our motivating application involves compositing a synthetic image with a natural image, our focus is on the illumination-related artifacts in such composites and not on any geometric errors introduced during the process. To avoid such errors, we generate stimuli from webcam images by compositing two images taken from the same webcam at different times of day.

3.1 Image Selection

In order to isolate the lighting changes due to the sun position, we first manually selected 1 day for each webcam sequence where the sky was completely clear. On each day, the sun moves through a different arc in the sky. As a result, the elevation and azimuth angles (as defined in previous work on lighting direction) are different for each point on the arc, but they vary monotonically.

Since webcam images are downloaded every 15 minutes or so, we select a subset of images based on their sun elevations, normalized by the maximum sun elevation that day, $\theta_s' = \theta_s / \max(\theta_s)$. This is done to account for variations in maximum sun elevations, which depend on the latitude of the camera and the time of year. In particular, we choose four images such that $\theta_s' = \{0.2, 0.4, 0.6, 0.8\}$, representing roughly early morning, late morning, early afternoon, and late afternoon. Finally, we also identify one large building in the scene, and manually segment it to obtain a mask M that is used to generate the composites.

3.2 Image Compositing

Given two images I_F and I_B, each lit with different lighting conditions F and B, and segmentation mask M, we generate the composite I_C as:

$$I_C = M \circ I_F + (1 - M) \circ I_B, \tag{1}$$

where $\circ$ denotes the Hadamard product. The resulting composite I_C, therefore, contains the foreground object of interest in lighting condition F and the background (including the shadow on the ground, if any) in lighting condition B.

Figure 1 shows example images of the webcam "Formica" from Lalonde et al. [2009] generated with this method. All possible combinations of the 4 lighting conditions from above are generated, resulting in a set of 16 images for each sequence. Images along the diagonal of the figure correspond to the cases where lighting between background and foreground are perfectly consistent.

3.3 Experiment I Design

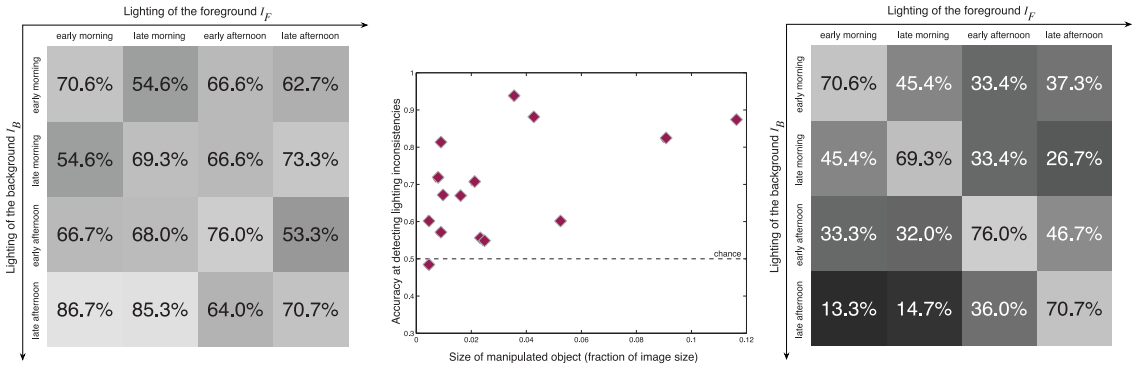
In this first experiment with five participants, we presented one image at a time to each observer, and asked them to say whether the image appeared real or fake. In all, we showed each of the 4×4 lighting combinations individually for each of the 15 scenes, for a total of 240 images, to each observer. Users had a maximum of 10 seconds to make a decision. After the 10-second delay was over, the images disappeared from view. Participants were told ahead of time which object was being potentially manipulated by showing the segmentation mask prior to displaying the image, but they were not told what kind of manipulations were being made.

3.4 Experiment I Results

Overall, the mean accuracy at identifying real versus manipulated was 68.08%, with chance being 50%. As shown in Figure 2(a), the particular set of lighting conditions did not seem to affect accuracy when averaged over all 15 scenes. We observe a slight increase in accuracy when the background is lit by late afternoon skies and the object by early to late morning conditions (bottom left of Figure 2(a)). Otherwise, performance across lighting configurations is relatively constant. We also analyze performance as a function of the size of the object in the image and show the results in Figure 2(b). As expected, illumination inconsistencies tend to be more perceptible in larger objects.

Figure 3 shows the accuracy in detecting inconsistencies for different scenes. All the results shown have the same illumination configuration; that is, the object is illuminated by early afternoon lighting, while the background was captured in the late afternoon. Even if the lighting configuration is the same, the average accuracy differs quite significantly across scenes.

Fig. 1. Example images used in Experiment I for the "Formica" webcam sequence from Lalonde et al. [2009]. The horizontal axis shows changes in the lighting conditions of the foreground object I_F, and the vertical axis shows changes in the lighting conditions associated with the background I_B.



(a) (b) (c)

Fig. 2. (a) Lighting inconsistency detection accuracy from Experiment I, as a function of the lighting configurations for the foreground and background, averaged over all 15 scenes in the study. Note that the diagonal shows the percentage of correct images identified as such, and the other values denote accuracy of anomaly detection. (b) Detection accuracy by size of the manipulated object, shown as a fraction of image size (in pixels). Overall, we note that accuracy seems to be relatively constant over the different lighting configurations (a), and it seems to increase with object size (b), (c) the accuracy results recast as percentage accepted as unmanipulated, with the drop off in values away from the diagonal showing the impact of lighting inconsistencies.

0% 20% 20% 60% 100%

Fig. 3. The accuracy in detecting lighting inconsistencies widely varies across scenes. Here, 0% means that no participant identified the image on the left as inconsistent, while all participants did so for the image on the right. In all these cases, the relative difference between the lighting conditions of the background and that of the object of interest are the same, that is: the object is early afternoon, and the background late afternoon. Yet, can you tell that all these images have been manipulated?

Table I. Description of the Metrics Used in the Histogram Comparison Experiments

Metric name	M_F	M_B
All	entire image	entire image
Bg–bg	background	background
Fg–fg	foreground	foreground
Fg–bg	foreground	background
Fg–boundary	foreground	constant-width border (5% image width) around foreground

3.5 Experiment I Histogram Analysis

In order to determine whether simple image metrics can predict how likely it is that a discrepancy will be detected, we compute image intensity histogram differences as follows. First, we convert the images I_F and I_B (from Section 3.2) into grayscale, and define masks M_F and M_B selecting different regions of each image. Then, we compute the difference in intensity distributions D between these two image regions by taking:

$$D = \chi^2(h(I_F(M_F)), h(I_B(M_B))), \tag{2}$$

where $h(\cdot)$ is the histogram of image intensities, and $I(M)$ represents the pixels in image I that lie under the mask M. We use the well-known χ^2 distance between two histograms. Table I provides an overview of the values for M used in the metrics we experimented with.

We calculated the correlation coefficients between the proportion of participants who detected each discrepancy (excluding the correct images in this case) and each of the metrics. We found significant correlations for Fg–fg, Fg-boundary and Fg-bg, although the coefficients were low ($0.31, 0.25, 0.20$, respectively). Nevertheless, there is some evidence to suggest that such metrics may explain some, but not all, results. However, it is clear that there are several other factors at play and that a more controlled experiment may help to elucidate these.

3.6 Experiment I Discussion

The results shown in Figure 2(a) appear to confirm the observation by Ostrovsky et al. [2005] that lighting inconsistencies are difficult to accurately detect. It might be expected that this would translate to being accepting of a wide range of images as unmodified. The diagonal of Figure 2(a) should have been 100% if observers were able to accurately identify consistent illumination conditions. The low values on the diagonal suggest that showing the images individually in the experiment does not purely assess the impact of the error in lighting direction, but also other judgments about the object and scene.

The entries in Figure 2(a) though are different from chance (50%), and light consistency appears to have an impact on the observer's judgment of whether the image has been manipulated. This effect is clear when the overall results are recast as the percentage of images accepted as unmanipulated in Figure 2(c), where the off-diagonal elements are sharply different from the diagonal. However the highly varying results for individual scenes portrayed in Figure 3 suggest though that scene structure may have an influence, for example, whether the element being composited is set apart from other parts of the scene or is closer to other elements in position and orientation. This observation leads us to the possibility that the answers to the questions we posted in Section 2 will be conditional on the type of scene.

Unfortunately, experimenting with real webcam data offers significant challenges in further pursuing answers to the questions we have posed. First and foremost, illumination cannot be controlled. Even despite the use of calibrated webcam data, we can choose only among a relatively limited set of light directions, which depend on the camera latitude and orientation. Second, the sun color and intensity both change over the course of the day. Finally, scenes cannot be controlled, so many factors such as scale, orientation, and viewpoint may vary from one scene to the next. These insights and limitations of the first experiment thus led us to the design of a second experiment that provided greater experimental control.

4. EXPERIMENT II: PHOTOREAL SYNTHETIC SCENES

To explore the impact of illumination when adding a new structure to a scene, we need to have control over both illumination conditions and scene geometry. However, while illumination may be easily manipulated in a lab setting, it is impossible to fully control the lighting conditions for real outdoor imagery, which we also learned from Experiment 1. For this reason, we chose to use a highly detailed digital model of an outdoor urban district, and render it under realistic outdoor lighting conditions.

4.1 Scenes

We used a model of Trinity College Dublin and surrounding streets that was created manually by artists in Autodesk 3D Studio Max [O'Sullivan and Ennis 2011]. Special attention was paid to accuracy and a high level of detail was incorporated via photographic texture maps. This model has been used in a variety of previous studies involving realistic urban scenes (e.g., Hamill et al. [2005] and Ennis et al. [2008, 2011]).

Based on our observations from Experiment I, we wanted to study two scenes with fundamentally different structures. We chose to focus on two complementary cases: (1) when an object is near but separated from existing structures and (2) when an object is in close contact with existing structures. The rich Dublin model offered several possibilities for both of these cases. The Campanile is a landmark on the Trinity College Dublin campus, and is isolated from the surrounding buildings in our first scene. It casts a shadow on the nearby ground plane that can easily be observed independently of its appearance. However, the overall texture and design style of the Campanile are similar to the other buildings in the scene. In contrast, the red brick building selected as the object of interest in our second scene is embedded with its neighbors on a street as one continuous facade, but it has a different texture and design style than its surroundings.

For each scene, we generated a subpixel accurate segmentation mask by rendering the selected object in white and the rest of the scene in black (disabling shading in the rendering engine). Figure 4 shows the scenes we selected (top row) and the corresponding objects of interest and background we selected (bottom row).

(a) Campanile (b) Street

Fig. 4. Overview of the two selected scenes used in Experiment II. The first row shows the original images, while the second row shows the segmentation masks (green = object, red = background).

4.2 Realistic, Controllable Outdoor Lighting

As mentioned previously, we employ a rendering approach because real imagery does not afford control over the natural lighting conditions. Our goal is to employ as realistic and natural an illumination model as possible. For this purpose, we employ the recently proposed parametric model of Lalonde and Matthews [2014], which represents the light intensity along direction $\mathbf{l}$ as a weighted sum of sky and sun components f_{sky} and f_{sun}, respectively:

$$f(\mathbf{l}) = \boldsymbol{\omega}_{\text{sky}} f_{\text{sky}}(\mathbf{l}, \mathbf{l}_s, t) + \boldsymbol{\omega}_{\text{sun}} f_{\text{sun}}(\mathbf{l}, \mathbf{l}_s, \beta, \kappa), \tag{3}$$

where $\mathbf{l}_s = (\theta_s, \phi_s)$ is the sun position, t is the sky turbidity, (β, κ) are parameters that control the shape of the angular scattering close to the sun, $(\boldsymbol{\omega}_{\text{sky}}, \boldsymbol{\omega}_{\text{sun}})$ are the sky and sun mean colors, respectively, f_{sky} is the well-known Preetham sky model [Preetham et al. 1999], and f_{sun} is a sun-specific model introduced in Lalonde and Matthews [2014]. We refer the reader to their paper for more details. Given values for its parameters, Equation (3) can be used to generate a hemispherical environment map (where only values above the horizon are valid) representing a physically plausible, high-dynamic-range (HDR) sky.

We obtain photo-realistic values for the sky model parameters by using the database of HDR natural skies also introduced in the same paper. The database contains HDR hemispherical photographs of the sky, captured over more than 3,300 different illumination conditions, on 25 different days, over the course of 6 months. The authors have also provided the results of fitting Equation (3) to each of their sky photographs, thereby resulting in a database of 3,300+ lighting parameters.

We cluster these lighting parameters into 10 subclusters using k-means, and keep the cluster that has the highest sun intensity, thereby modeling a clear day. Since the mean sun color $\boldsymbol{\omega}_{sun}$ is much brighter than any of the other parameters (values may go up to 10^4), we use its log-value in the clustering algorithm. Figure 5 illustrates the resulting environment obtained with this method, shown

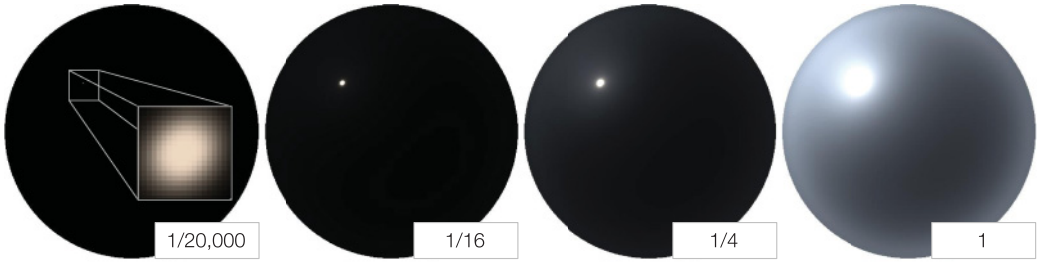


Fig. 5. HDR lighting model used to relight the virtual scene, shown at different exposures. The numbers are fractions of the original exposure, shown on the right. The sun is not saturated only when dividing the original exposure by 20,000, which shows the extreme dynamic range represented by our lighting model. Note that clouds present in the original data, but not captured by Lalonde's parametric model of 2014, make the mean sky color tend toward gray.

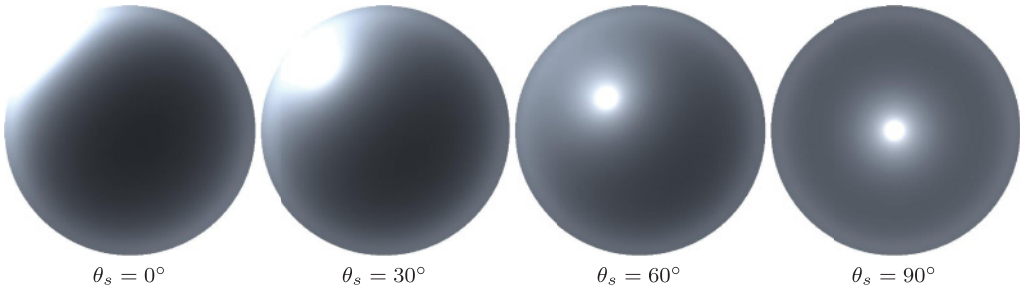


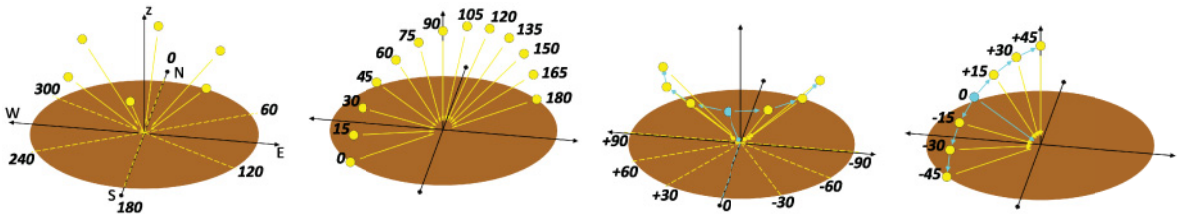
Fig. 6. Effect of varying the elevation of the sun θ_s in our lighting model (ϕ_s is kept constant at 45°). Note: capturing a sky with the sun directly at zenith ($\theta_s = 90°$) would be challenging because this only happens between the tropics at certain times of the year. With our model, this can easily be replicated. Also note that we do not change the sun intensity as a function of elevation.

at different exposures to reveal its extremely high dynamic range. Since this is a parametric model, we can choose to place the sun at any position in the hemisphere, and obtain a physically plausible sky, as demonstrated in Figure 6. Also note how the resulting sky is smooth because Equation (3) does not model clouds. This is acceptable in our case, since clouds could act as a distraction to users. All the images used in our experiment were rendered with the resulting environment map as the sole light source.

4.3 Image Generation

From the lighting environment obtained in the previous section, we rendered images of both scenes using the MentalRay renderer in 3D Studio Max, by enabling soft shadows and final gather. We set the environment map as the sole light source. For the Campanile scene, hair and fur modifiers (WSM) were used to achieve more realistic grass rendering. Note that we use the renderer to generate images where the lighting is perfectly "consistent" and generate "inconsistent" images by combining two real images that have different lighting directions with a compositing approach. Here we describe the lighting directions employed to generate the environment maps from Section 4.2. Images were composited using the same procedure as in Section 3.

4.3.1 *Azimuth Increment.* To limit the test cases to a reasonable size but ensure a sufficient coverage of the azimuth circle, we define a set of foreground angles with increments of 60°: $\phi_s = [0, 60, 120, 180, 240, 300]°$. Elevation is kept constant at $\theta_s = 45°$ (see Figure 7(a)). From these foreground angles,



(a) Foreground ϕ_s (b) Foreground θ_s (c) Background $\Delta\phi_s$ (d) Background $\Delta\theta_s$

Fig. 7. Foreground and background sun azimuth and elevation samples. In these drawings, the camera is looking straight north. Angles are shown for (a) foreground azimuth ϕ_s, in increments of 60° (with $\theta_s = 45°$), (b) foreground elevation θ_s, in increments of 15° (with $\phi_s = [45, 135]°$), (c) background azimuth, in relative increments of 30° from the foreground angle, and (d) background elevation, in relative increments of 15°. In both (c) and (d), the foreground angle is drawn in cyan.

we define relative changes by increments of 30° away from the original angle: $\Delta\phi_s = [-90, -60, -30, +30, +60, +90]°$ to obtain background angles. This approach is similar to the one use in the work of Lopez-Moreno et al. [2010].

4.3.2 *Elevation Increment.* In addition to azimuth, we also test for sensitivity to sun elevation variations. We selected foreground elevation angles by 15° increments, $\theta_s = [0, 15, 30, 45, 60, 75, 90]°$, as shown in Figure 7(a). Azimuth is kept constant at $\phi_s = [45, 135]°$ (two diametrically opposed values). From these foreground angles, we define relative changes by increments of 15° away from the original angle: $\Delta\theta_s = [-45, -30, -15, +15, +30, +45]°$ to obtain background angles. Figure 5(c) shows an example of relative changes shown for a given foreground angle.

As opposed to the azimuth-only case, we impose additional constraints on the generation of composite images in the case of elevation. In particular, we do not include the case where $\theta_s + \Delta\theta_s < 0°$, which corresponds to the sun below the horizon. In addition, we discard the case where $\theta_s + \Delta\theta_s \geq 90°$, which means that the sun moves past zenith, and down the opposite direction.

4.4 User Study Design

Fourteen naïve volunteers (9M/5F, ages 24–36) participated in this experiment. We did not wish to ask participants to judge each image individually, as we found in Experiment I that the nature of each scene could influence them. Therefore, we designed a comparison task using an experimental procedure described by Hospedales and Vijayakumar [2009] for oddity detection. Rather than simply asking the observer which of two stimuli is more realistic, we present two correct scenes alongside one altered scene and ask the observer to select the least realistic one. Even if we are modifying only the sun direction, observers may rely on many different cues which may point to illumination inconsistencies in an image. For example, cast shadows, shading, and overall brightness may all be good reasons for deciding whether an image composite is consistent. Since we do not want to bias our participants toward preferring one cue over another, we created a simple user interface in which the three images are arranged horizontally on a straight line and provided a radio button interface that forces the user to make a selection before the next image can be displayed. Images are shown in randomized order. Each fake image was tested three times in different sets of three images. Users had a maximum of 8 seconds to view the images, after which they disappeared.

5. RESULTS

We analyze the results of Experiment II for variations in the sun azimuth and elevation independently.

Table II. Main Significant Results for the Experiments Presented in This Article

AZIMUTH

Effect	F-Test	Post-hoc
DIR	$F_{1,13} = 5.84$, $p < 0.05$	Changes to right of camera higher
SIDE	$F_{1,13} = 81.01$, $p \approx 0.05$	Front > Back
ERR	$F_{2,26} = 28.65$, $p \approx 0.00$	$30 < 60 < 90$
AZI	$F_{2,26} = 16.63$, $p \approx 0.00$	Highest when lit from center or left of camera
MODEL×DIR	$F_{1,13} = 5.94$, $p < 0.05$	Street: Left<Right
MODEL×AZI	$F_{2,26} = 7.72$, $p < 0.005$	Left of camera lower for Street only
DIR×AZI	$F_{2,26} = 4.27$, $p < 0.05$	Left dir vs. Right not sig. when lit from right
MODEL×SIDE×ERR	$F_{2,26} = 8.50$, $p < 0.005$	Order of front and back errors different for Street and Campanile.
MODEL×DIR×AZI	$F_{2,26} = 29.28$, $p \approx 0.00$	Biggest change is in Street, lit from left, changedir to left
MODEL×SIDE×AZI	$F_{2,26} = 5.45$, $p < 0.05$	Street: lit from left and back lower than from front. Not so Campanile
MODEL×ERR×AZI	$F_{4,52} = 4.93$, $p < 0.005$	Best explained in five-way interaction
DIR×SIDE×ERR	$F_{2,26} = 11.55$, $p \approx 0.00$	See five-way interaction
DIR×SIDE×AZI	$F_{2,26} = 10.84$, $p \approx 0.00$	See five-way interaction
MODEL×SIDE×ERR×AZI	$F_{4,52} = 9.72$, $p \approx 0.00$	See five-way interaction
DIR×SIDE×ERR×AZI	$F_{4,52} = 5.86$, $p < 0.005$	See five-way interaction
MODEL×DIR×SIDE×ERR×AZI	$F_{4,52} = 6.41$, $p \approx 0.00$	See Figure 8 for full details

ELEVATION

Effect	F-Test	Post-hoc
ELE	$F_{7,91} = 5.33$, $p \approx 0.0$	Zenith, Horizon and midpoints higher
ERR	$F_{2,26} = 15.62$, $p \approx 0.0$	$15 = 30 < 45$
MODEL×SIDE	$F_{1,13} = 12.71$, $p < 0.005$	Front < Back for Campanile, opposite for Street
MODEL×ERR	$F_{2,26} = 8.16$, $p < 0.005$	$15 = 30 < 45$ for Campanile, all diff for Street
ELE×ERR	$F_{14,182} = 2.69$, $p < 0.005$	See 3-way interactions
MODEL×SIDE×ERR	$F_{2,26} = 3.76$, $p < 0.05$	Campanile: Front > Back only for 45; Opposite for all Street ERR.
MODEL×SIDE×ELE	$F_{7,91} = 3.45$, $p < 0.005$	See Figure 9 for full details

5.1. Azimuth

We ran a five-way, repeated measures analysis of variance (ANOVA) on the results of the Azimuth data, with factors *Model* (Campanile, Street), *ForegroundAzimuth* (the *foreground* sun azimuth angle from Section 4), *Side* (whether the foreground is lit from the front or back), *ChangeDirection* (sun going toward the left of the camera, or toward the right), and *Error* (differences in the background azimuth angle of $[30, 60, 90]^\circ$.). We performed post-hoc analysis using Newman-Keuls tests of differences between means to examine the effects further. The results are shown in Table II and Figure 8.

There were significant interaction effects between almost all combinations of factors, which is an indication of how many different types of changes to the positioning of the sun, and the surrounding environment, can result in a different perception of the illumination of the scene. Note that in both of the following graphs, a random guess would result in 33% performance, since we show three images at a time (c.f. Section 4.4). Therefore, effects that proved to be significantly different from chance as indicated by single sample t-tests are shown in bold color.

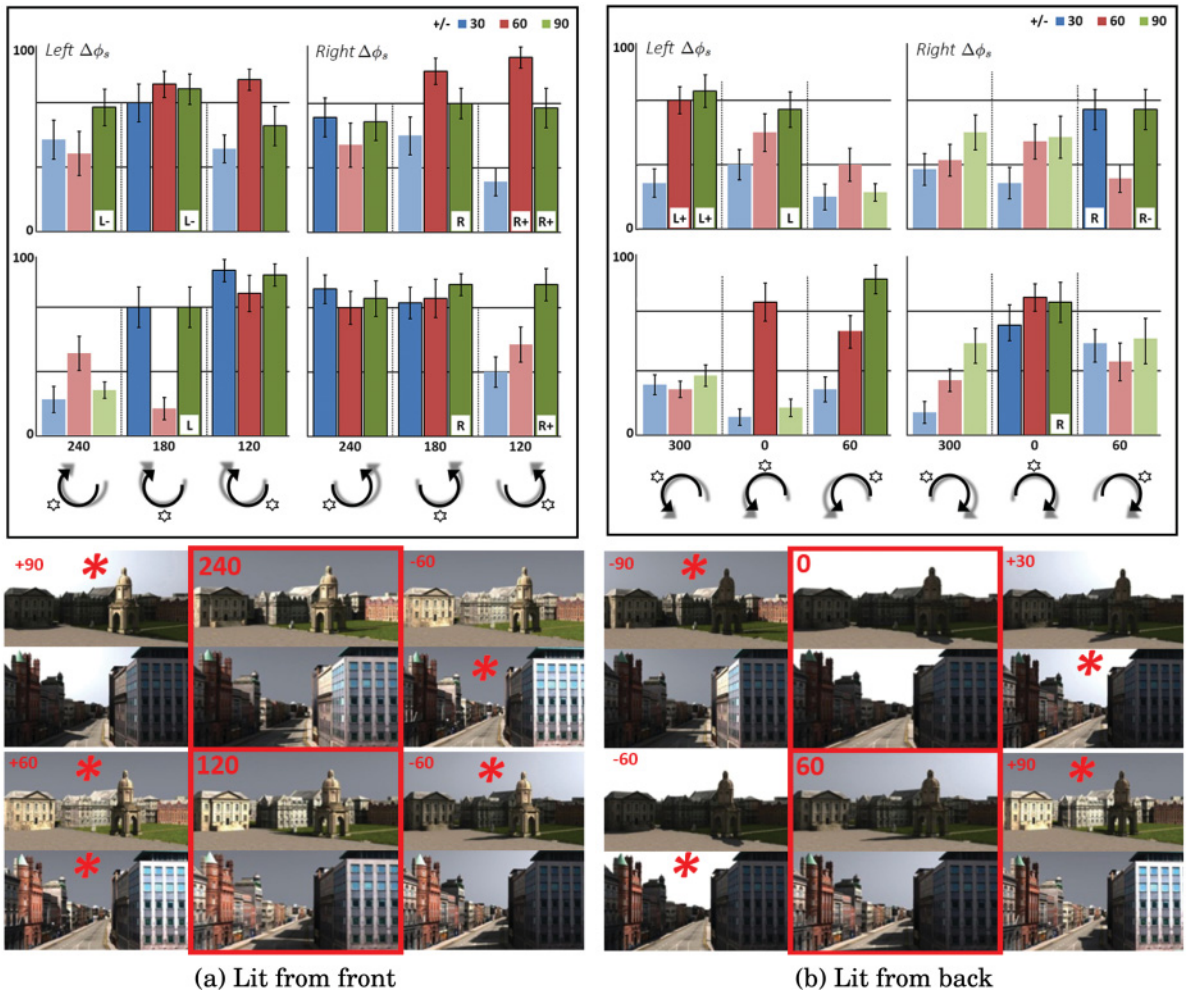


Fig. 8. Results for sun azimuth ϕ_s changes, when the sun is (a) in front of, or (b) to the back of the scene (top Campanile, bottom Street in all graphs). Error bars show standard errors and R(+), L(-) indicate whether the scene boundary was reached or crossed. The arrows show the direction of the Δ error and the icons/dots and axis labels show the foreground sun azimuth. Sample images with perceptible differences are marked with a red asterisk (*). The images with red outlines have consistent illumination, and other images correspond to $\Delta\phi_s = \pm[30,\ 60,\ 90]^\circ$.

First, participants were indeed able to distinguish all error sizes under some circumstances. The most significant effect to note is that when the Foreground Azimuth was in front of the scene, that is, at 180°, the errors were most noticeable for both models. Figure 8 also shows several examples of images that were shown to the participants of our study.

ChangeDirection also had a significant effect, where shifting the sun to the right of the scene (with respect to the camera viewpoint) was more noticeable than shifting to the left. When the sun reached or crossed the left or right boundaries of the scene, this was also quite noticeable in some, but not all, cases. There is an important discrepancy between the two scenes when $\phi_s = 0^\circ$: We believe this asymmetry was due to the particular geometry of the street scene. In this scene, the object of interest has one main plane, facing toward the right. Therefore, when the sun is at 0° (directly in front of the

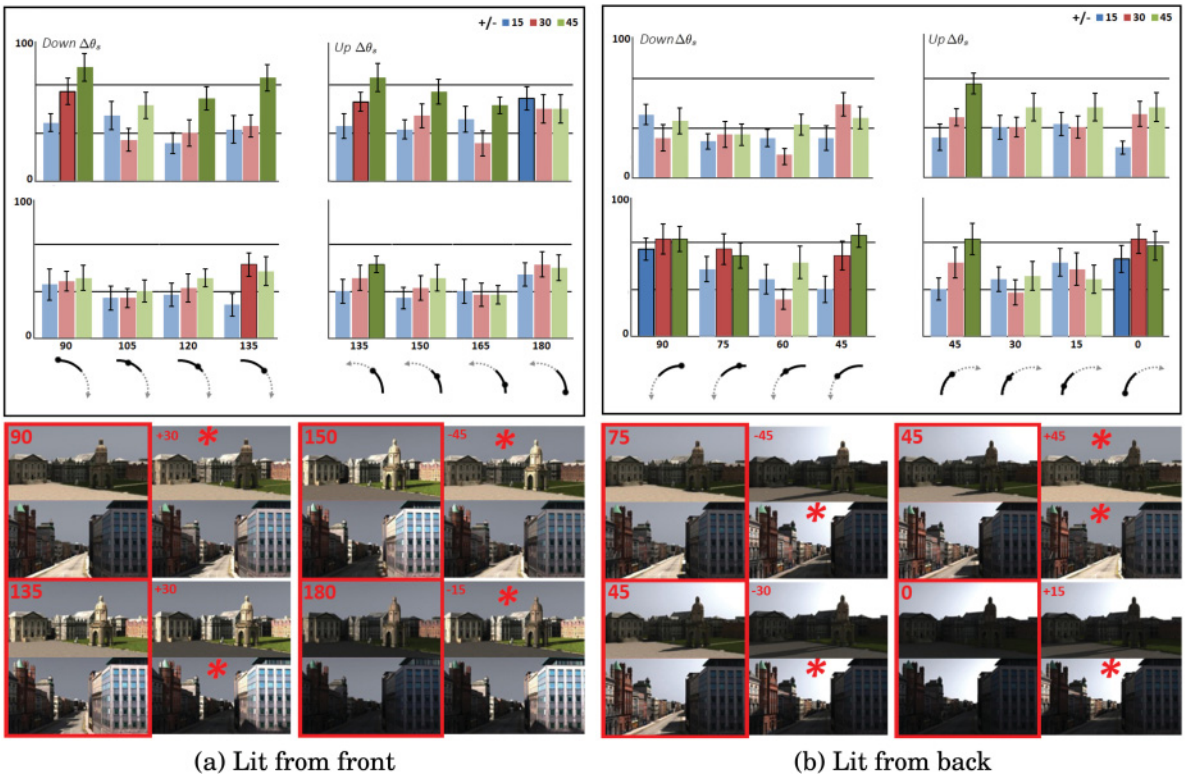


Fig. 9. Results for sun elevation θ_s changes, when the sun is (a) in front of, or (b) to the back of the scene (top Campanile, bottom Street in all graphs). Error bars show standard errors and R(+), L(-) indicate whether the scene boundary was reached or crossed. The arrows show the direction of the Δ error and the icons/dots and axis labels show the foreground sun elevation. Sample images with perceptible differences are marked with a red asterisk (*). The images with red outlines have consistent illumination, and other images correspond to $\Delta\theta_s = \pm[15,\ 30,\ 45]^\circ$ as indicated.

camera) and it moves to the *right*, it shines straight on the building, thereby making changes more noticeable. When it moves to the *left*, however, it shines on the back of the building, so the effects are less visible from the camera viewpoint.

5.2 Elevation

We also ran a four-way, repeated measures ANOVA on the results of the Elevation data, with factors *Model* (Campanile, Street), *ForegroundElevation* (the foreground sun elevation angle from Section 4), *Side* (whether the foreground is lit from the front or back), and *Error* (differences in the background azimuth angle of $[15, 30, 45]^\circ$.) We also tested whether the direction of the change (up or down) had an effect, and found that it did not, so we removed that factor from our analysis (though it is shown in the graph for completeness). We again performed post-hoc analysis using Newman-Keuls tests of differences between means to examine the effects further. The results can be seen in Table II and Figure 9.

It should be noted that the size of the errors in the Elevation test were smaller (15,30,45), as we only have one hemisphere to work within, and we wished to have an equal number of observations for each condition. Taking into account the smaller error sizes, there are fewer noticeable changes than with the Azimuth condition. However, again we notice some interesting differences depending on the

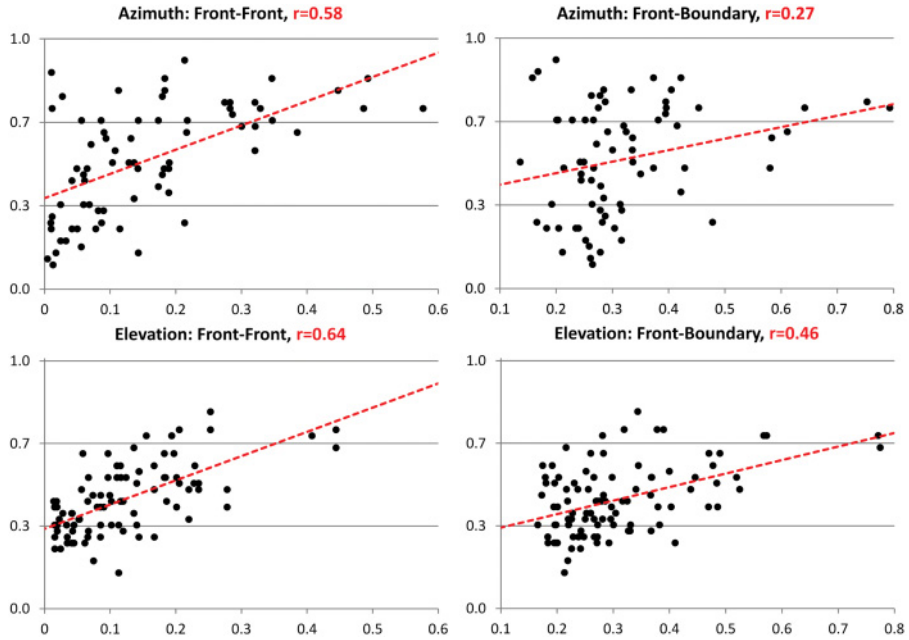


Fig. 10. Correlations between participant detection accuracy and the two most promising image histogram metrics.

context. For example, when the Campanile is lit from the front, elevation changes are quite noticeable, whereas from the back, only one bigger change was detected above chance. However, the opposite is true for the Street scene, where changes were most noticeable when the sun was lit from the back on the right. This is because of the particular nature of the street scene, where the sun shines from the right on the red building.

5.3 Histogram Analysis

We compute the same histogram metrics as described in Section 3.5 and the most significant correlations are shown in Figure 10. The most interesting observations we made are as follows:

Azimuth: there is a significant correlation between the front-front histogram metric values and participant accuracy in detecting inconsistencies, which is higher than any other metric we tested. This is interesting because for the front-front metric, the participants do not actually see what the object should have been. This suggest that they may be using some higher-level understanding of the scene to infer what the object should have looked like, given the lighting conditions dictated by the background.

After further analysis of the data, we found that the front-boundary metric does not actually predict detection accuracy for the Campanile scenes, whereas it does do so for the Street scene. This seems to suggest that the geometric relationship between the object and its surroundings is important: when the object is detached (Campanile), then local intensity changes are not sufficient to predict performance; however, when the object is attached to its surroundings, and thus shares geometric properties such as main orientation (Street), then local intensity changes do predict performance more accurately.

We also found that the correlation with the front-front metric is strongest when the scene is lit from the back, which seems to suggest that it is easiest to detect inconsistencies when the object is dark and the background bright rather than vice versa. Additional studies are needed to explore this result further.

Elevation: as with the Azimuth histogram data, the strongest correlation is observed with the front-front histogram metric in the case of elevation. However, we also observed more correlations with the other metrics, which seems to suggest that global intensity changes are more important in the case of elevation.

6. DISCUSSION

6.1 Insights on Original Questions

From the results of Experiment II, we can suggest some answers to the questions posed in Section 2:

(1) *Can we quantify the noticeable light source inconsistencies for sun direction relative to view direction in outdoor scenes?*
We cannot put any absolute bounds on whether a particular inconsistency is noticeable. For both azimuth and elevation, there were cases where the smallest change we introduced was perceptible and the largest case was not. Consistent with previous work [Koenderink et al. 2003], observers were less sensitive to changes in elevation. We note that this adds to our understanding because in previous work the "camera view" was near 90° elevation, rather than 0° elevation (closer to normal street view) in our study. Unlike previous work, we found that it was not just the magnitude of the error but the absolute value of the light direction that mattered.

(2) *Do we see the same thresholds for angle as observed by Lopez-Moreno et al. [2010] for the real and textured scenarios? Do we see the same backlit/frontlit effects?*
Despite the presence of texture in our study, we found in several cases that errors of 30° were easily perceptible. We found differences in frontlit and backlit scenes, but we did not find that illumination inconsistencies from the back were more readily detected. Our results were dependent on scene structure. In particular, we are considering inserting building objects with strong orientation in a particular direction, rather than the rounded objects considered by Lopez-Moreno et al. As a result, the errors in the light direction with respect to the primary orientation of the object being inserted have a greater effect.

(3) *Do we see the same thresholds for angles vary with respect to height above the viewer (elevation) as for angular variations from left to right (azimuth)?*
We considered differences of 30° for both azimuth and elevation. While differences of 30° in azimuth were perceptible in a wider range of circumstances, differences in 30° of elevation were important for either (but not both) front illumination or back illumination, depending on the scene structure.

(4) *Is the sensitivity for large structures in large depth-of-field images the same as for sets of synthetic or low-depth-of-field objects?*
The sensitivity for large structures is different, apparently not because of texture masking or "naturalness," but because of the strong consistent orientation of the objects we are considering.

(5) *Is there an effect of whether light comes from the left or right of the viewer?*
The effect of light coming from the left or right of the viewer is different for the two different scenes. It appears that the orientation of the light with respect to the main orientation of the element of interest is more important than light direction with respect to the viewer.

6.2 Application: Determining the Realism of Previsualization Composites

We return to the original motivation for our work and apply our findings to the application of determining the realism of previsualization image composites, as are commonly needed in architectural design or film production.

Fig. 11. Examples of real previsualization image composites. We show the composite images on the top row, and their corresponding segmentation masks bottom row for reference. The masks indicate the inserted object (green) and the original background image (red). In both images, there is a difference of 30° in sun azimuth between the inserted object and the background. According to our findings, users are typically *not* able to tell whether there is an illumination inconsistency on the image on the left, but they *are* for the image on the right. Our results could be helpful in designing a previsualization system that takes illumination inconsistencies into account.

For the purpose of demonstration, we created image composites manually from images where the relative position of the sun with respect to the camera is known. In Figure 11, we generated two montages by transferring a structure from one image onto another. Montage 1 was created starting from an image where the sun had an azimuth angle of $\phi_s = 120°$ (following the notation in Figure 7(a)), whereas Montage 2 was created from an image where $\phi_s = 180°$.

In both cases, we inserted a structure cut from an image with a 30° azimuth difference from the background lighting conditions. All the images used to generate montages had a sun elevation of approximately $\theta_s = 45°$. This situation corresponds to the case shown in Figure 8(a). Indeed, we notice that, while montage 1 looks correct, there is a perceivable illumination inconsistency in Montage 2: the building on the left has improper shading with respect to the one on the right. Even if the azimuth difference is constant at 30° in both cases, the effect of the *absolute* location of the main light direction is important.

7. CONCLUSIONS

In this article, we investigate the conditions in which lighting inconsistencies in montages of outdoor scenes are perceivable by humans. Focusing on the directionality of outdoor lighting, we present findings gathered from a user study asking participants to identify inconsistent composites generated from a database of synthetically rendered images.

Our first main contribution is to show that previous studies done on simpler, toy-like stimuli, corroborate most of our findings in the context of outdoor scenes. In particular, these studies have demonstrated that humans seem to be insensitive to directional lighting inconsistencies when azimuth changes are below 40° [Lopez-Moreno et al. 2010], and that changes in elevation are even harder to notice [Koenderink et al. 2004]. Our second main contribution is a novel observation, not mentioned by previous studies, which is that when the sun is directly behind the camera, the threshold at which observers spot illumination inconsistencies seems to *decrease*, and it becomes harder to generate a realistic composite. Our findings are useful to guide the generation of previsualizations for outdoor scenes.

Our approach has the following limitations. First, in order to reduce the number of test cases to a reasonable minimum, we sampled the space of azimuth angles at the relatively coarse increments of 60°, and relative changes at 30° jumps. It would be interesting to repeat a similar study by reducing this increment (especially the *relative* increment) to a lower value, to determine if a more precise threshold can be recovered. Second, the set of scenes used in the study was limited to two, again to keep the total number of conditions to evaluate to a manageable number. The Dublin model presented in Section 4.1 is very rich, and could easily be used to generate a more varied set of scenes. Third, objective measures such as the light directionality are important, but probably not sufficient. Future work should explore additional lighting properties such as intensity and chromaticity, which are indissociable from the sun position in real-world imagery.

ACKNOWLEDGMENTS

The authors would like to thank Alyosha Efros for fruitful discussions in the early stages of this work.

REFERENCES

Patrick Cavanagh. 2005. The artist as neuroscientist. *Nature* 434, 301–307.

Xiaowu Chen, Ke Wang, and Xin Jin. 2011. Single image based illumination estimation for lighting virtual object in real scene. In *CAD Graphics 2011*. 450–455.

Cathy Ennis, Christopher Peters, and Carol O'Sullivan. 2008. Perceptual evaluation of position and orientation context rules for pedestrian formations. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV'08)*. 75–82.

Cathy Ennis, Christopher Peters, and Carol O'Sullivan. 2011. Perceptual effects of scene context and viewpoint for virtual pedestrian crowds. *ACM Transactions on Applied Perception (TAP)* 8, 2 (Feb. 2011), Article 10, 22 pages.

J. Hamill, R. McDonnell, S. Dobbyn, and C. O'Sullivan. 2005. Perceptual evaluation of impostor representations for virtual humans and buildings. *Computer Graphics Forum* 24, 3, 623–633.

Timothy Hospedales and Sethu Vijayakumar. 2009. Multisensory oddity detection as Bayesian inference. *PLoS ONE* 4, 1 (01 2009), e4205.

Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem. 2011. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (SIGGRAPH Asia'11)* 30, 6 (Dec. 2011), 157:1–157:12.

Kevin Karsch, Kalyan Sunkavalli, Nathan Carr, Hailin Jin, Rafael Fonte, Michael Sittig, and David Forsyth. 2014. Automatic scene inference for 3D object compositing. *ACM Transactions on Graphics (TOG)* 33, 3 (May 2014), 32:1–32:15.

Jan J. Koenderink, Andrea J. van Doorn, Astrid M. L. Kappers, Susan F. Te Pas, and Sylvia C. Pont. 2003. Illumination direction from texture shading. *JOSA A* 20, 6, 987–995.

Jan J. Koenderink, Andrea J. van Doorn, and Sylvia C. Pont. 2004. Light direction from shad(ow)ed random Gaussian surfaces. *Perception* 33, 12, 1405–1420.

Jean-Francois Lalonde and Alexei A. Efros. 2007. Using color compatibility for assessing image realism. In *IEEE International Conference on Computer Vision (ICCV'07)*. 1–8.

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. 2009. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. *ACM Transactions on Graphics (SIGGRAPH Asia'09)* 28, 5 (Dec. 2009), 131:1–131:10.

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. 2012. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision* 98, 2, 123–145.

Jean-François Lalonde and Iain Matthews. 2014. Lighting estimation in outdoor image collections. In *Proc. of the International Conference on 3-D Vision (3DV'14)*. 131–138.

Jorge Lopez-Moreno, Elena Garces, Sunil Hadap, Erik Reinhard, and Diego Gutierrez. 2013. Multiple light source estimation in a single image. *Computer Graphics Forum* 32, 8, 170–182.

Jorge Lopez-Moreno, Veronica Sundstedt, Francisco Sangorrin, and Diego Gutierrez. 2010. Measuring the perception of light inconsistencies. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization (APGV'10)*. ACM, New York, NY, 25–32.

Yuri Ostrovsky, Patrick Cavanagh, and Pawan Sinha. 2005. Perceiving illumination inconsistencies in scenes. *Perception* 34 (Nov. 2005), 1301–1314.

Carol O'Sullivan and Cathy Ennis. 2011. Metropolis: Multisensory simulation of a populated city. In *Proceedings of the 2011 3rd International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES'11)*. 1–7.

A. J. Preetham, Peter Shirley, and Brian Smits. 1999. A practical analytic model for daylight. In *Proceedings of ACM SIGGRAPH 1999*. 91–100.

Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. 2001. Color transfer between images. *IEEE Computer Graphics and Applications, special issue on Applied Perception* 21, 5 (Sept. 2001), 34–41.

Su Xue, Aseem Agarwala, Julie Dorsey, and Holly Rushmeier. 2012. Understanding and improving the realism of image composites. *ACM Transactions on Graphics (TOG)* 31, 4, 84:1–84:10.

Received July 2015; accepted July 2015