Computers & Graphics ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Compositing images through light source detection

Jorge Lopez-Moreno [a,*], Sunil Hadap [b], Erik Reinhard [c], Diego Gutierrez [a]

[a] Universidad de Zaragoza, Spain
[b] Adobe Systems Inc., United States
[c] University of Bristol, United Kingdom

ARTICLE INFO

Keywords:
Light detection
Image compositing
Relighting
Psychophysics

ABSTRACT

Compositing an image of an object into another image is a frequently occurring task in both image processing and augmented reality. To ensure a seamless composition, it is often necessary to infer the light conditions of the image to adjust the illumination of the inserted object. Here, we present a novel algorithm for multiple light detection that leverages the limitations of the human visual system (HVS) described in the literature and measured by our own psychophysical study. Finally, we show an application of our method to both image compositing and synthetic object insertion.

1. Introduction

This paper deals with the problem of obtaining the positions and relative intensities of light sources in a scene, given only a photograph as input. This is generally a difficult and under-constrained problem, even if only a single light source illuminates the depicted environment.

Traditionally, light probes are used to acquire the lighting data [1,2]. A light probe is an object of known 3D shape and BRDF properties (Bidirectional Reflectance Distribution Function, which is a description of the reflectance properties of the material) that is positioned in the scene when the image is captured. Unfortunately, in several cases, this technique is not applicable; e.g., in paintings and in photographs taken under uncontrolled conditions. It would be possible to use any object in the image if geometry information was available to allow light source positions or directions to be estimated [3,4]. Conversely, if the light source is known, the 3D geometry can be approximately recovered, an ill-posed problem known as shape-from-shading [5].

However, we are interested in the problem of light source recovery without the benefit of *any* geometric prior models. To this end, we first carried out a psychophysical experiment to quantify the accuracy with which humans can generally detect light sources. The results of this experiment were then used to validate the results of our light-detection algorithm, both numerically and perceptually. We then used any existing object in the image as a de facto light probe. We found that assuming a globally convex shape for such a light probe is sufficient to reconstruct light directions. The user only needs to identify the silhouette of the object in the image, a task similar to or simpler than other existing image-editing applications [6,7]. We then analyzed the information in the contour and the gradients contained in the shape to infer the light directions and relative intensities.

Real environments are likely to contain multiple light sources. In practice, we found that identifying up to four sources that when combined provided similar illumination as in the image sufficed for most situations. This keeps the dimensionality of the solution manageable, in a way similar to professionally lit environments, which are usually lit by a three-light setup. Additionally, although we did assume in principle that the chosen light probe was Lambertian, we will show that this is not a strong requirement.

We believe that by analyzing the lighting consistency between images, our algorithm can help improve several types of applications, such as Photo Clip Art [7], Interactive Digital Photomontage [8] or Photo Tourism [9].

2. Previous work

The computation of light source directions from images is an ill-posed problem, with many possible solutions leading to the same observed image. As a result, assumptions about the environment must be made, known geometry must be present in the scene, or extra information must be captured to change the problem into a solvable one.

To detect single light sources, a local analysis of the surface and image derivatives may be used to estimate the direction of the light source [10–12]. Alternatively, occluding contours within a single object [13,14] or texturing [15,16] provide clues as to where the light is coming from.

* Corresponding author. Tel.: +34 976 762 353.
E-mail address: jlm@unizar.es (J. Lopez-Moreno).

0097-8493/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2010.08.004

To detect multiple lights, the environment could be photographed along with the aforementioned light probe—a calibration object of known size and shape. For instance, a Lambertian sphere could be employed and subsequently analyzed for multiple light source directions [17,18]. Alternatively, multiple specular spheres can be triangulated for the same purpose [19,20]. Combinations of Lambertian and specular spheres have also been used [21]. Finally, reflections of a human eye can be analyzed to detect light sources [22].

Another way to overcome the underconstrained nature of the problem is to use a range camera to record geometry, allowing light sources to be inferred from the combination of the photograph and the range data [23]. Known geometry can be used to the same effect [24,25]. In contrast, our approach is free of previous restrictions; e.g., there is no need for a calibration object or known geometry. Furthermore, we do not require shadows to be cast on nearby objects, nor is any camera information needed.

3. Perceptual framework

Natural illumination in real environments is often complicated, making its analysis by both machines and humans difficult. Natural illumination exhibits statistical regularities that largely coincide with those found for images of natural environments [26,27]. In particular, the joint and marginal wavelet coefficient distributions, harmonic spectra, and directional derivative distributions are similar. Nonetheless, a complicating factor is that illumination is not statistically stationary because of locally dominant light sources [28]. By representing illumination with spherical harmonics, Mury et al. [29] have recently shown that low-order components show significant regularities, whereas the statistical non-stationarity is captured in the higher frequencies. Moreover, variation in the low-frequency representation tends to covary with the geometry rather than with the illumination. This interpretation is consistent with evidence suggesting that human vision assumes a priori the global convexity of object shapes [30]. Thus, human vision may apply the dark-is-deep paradigm, namely, that globally darker shading values indicate surface points that are further away than lighter values. Natural scenes, however, contain significant high-frequency components, and these complicate analysis. It is possible that human vision ignores these components, and this may help explain why human vision is not accurate in the perception of illumination in cluttered environments [31].

3.1. Psychophysical quantification

Ostrovsky et al. [31] show that even though the visual system can easily spot an anomalously lit object in an array of identical objects with the same orientation and lit in exactly the same way [32,33], the overall performance drops when altering orientations of the equally lit objects. This suggests that a fast, parallel pattern-matching mechanism in the former case is substituted with a much slower serial search in the latter, making illumination detection a difficult task for humans.

Because of uncontrolled factors such as lens distortion or glare in the input images, detecting light directions cannot be absolutely precise. We are therefore interested in determining an error threshold below which variations in the direction vector of the lights will not be noticed by a human observer. For this, we performed a psychophysical experiment inspired by [31,34,35]. Participants had to spot an anomalously lit item among a set of identical objects with randomized positions and orientations (see Fig. 1, bottom). We limited the test to the most restrictive azimuth angle ϕ (see Fig. 2d); it has been observed that human perception is better at azimuth estimates than at zenith θ estimates [36]. In their experiments, Koenderink et al. asked human observers to estimate the illumination direction for samples of random Gaussian surfaces illuminated by a collimated beam from random directions. The divergence between the coherent and the anomalous light in our test varied between 5 and 100°.

Eighteen participants took part in the experiment, none of them computer graphics experts. All reported normal or corrected-to-normal vision. Each was shown the entire sequence of images in random order and was asked to detect the inconsistently lit object. No time limits were imposed on the task. The results, shown in the chart

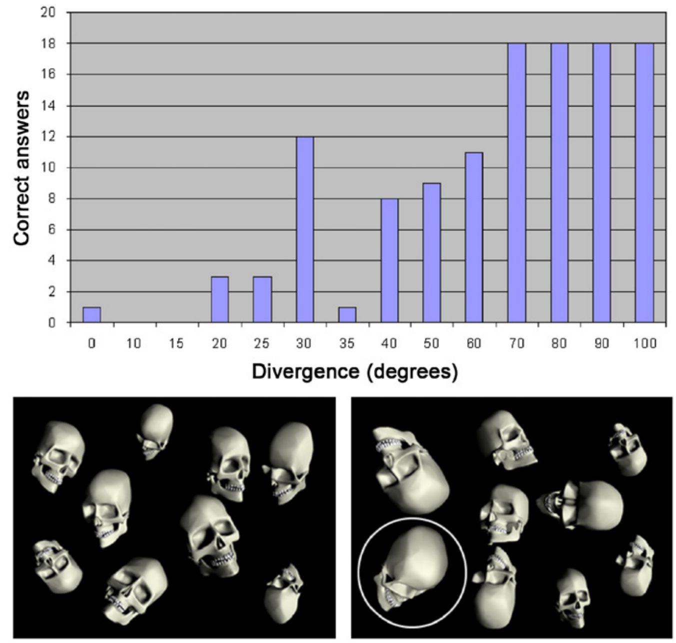

Fig. 1. Psychophysical test. Top: number of correct answers for each stimulus. Bottom: example stimuli shown to the participants. The rightmost image shows the anomalous object for the outlier (highlighted for visualization purposes).

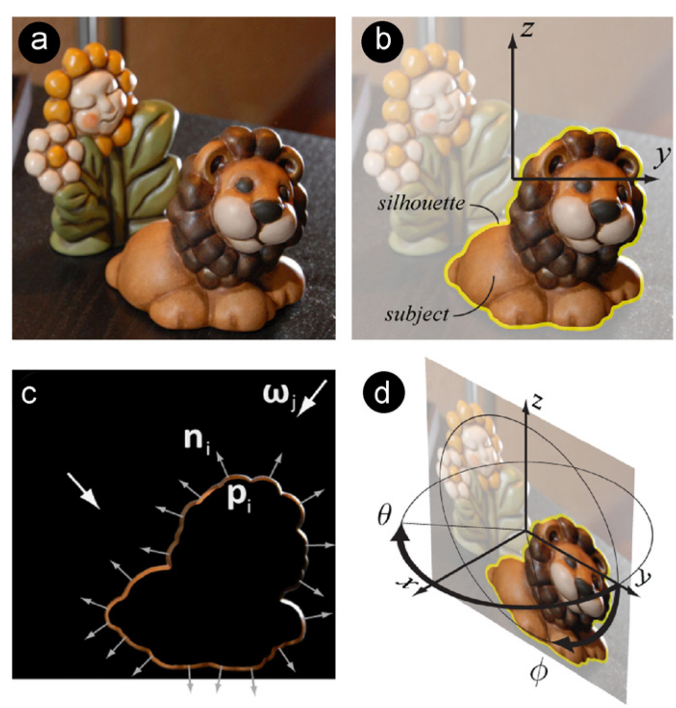

Fig. 2. (a) Input image, (b) object, (c) silhouette normals, and (d) coordinate system.

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

in Fig. 1, show how for up to 35°, the participants failed to guess correctly in most cases (only 3 right answers at best). An exception occurred at 30°, where the anomalous object was spotted by 12 participants. However, as the overall tendency confirms, this can be considered an outlier: the randomized position and orientation of the anomalous object combined in that specific image to provide obvious clues (see Fig. 1, bottom right). For divergences of 70° or more, all the participants guessed correctly. We thus assumed a conservative threshold at 35°, which is well above the measured error of our algorithm (see Section 5). Note that in our examples throughout this paper, we used images of different objects, so it seems safe to assume that the threshold would be even higher, given that the observer could not rely on direct comparison between identical objects.

4. Light detection

Consider a typical input image such as is depicted in Fig. 2a. The problem at hand was to estimate the number of illumination sources, their dominant directions and the relative intensities. We propose that any object in the image can be used as a virtual light-probe as long as it covers a reasonable area in the image. The user provides the outline defining the object, typically with the aid of a smart selection tool [37]. We do not assume any restrictions on the shape, the color or any other features of the object.

4.1. Assumptions and overview

To achieve a reasonable solution, we relied on the particular characteristics of human vision. In estimating illumination, the human visual system tends to ignore the local shape variations, and treats the object as a globally convex geometry [30]. We also leveraged the tendency of the human visual system to perceive objects correctly as long as the illumination is locally consistent [31]. Further, we observed that humans are surprisingly good at estimating back-lighting using cues from shadows [36]. Based on these assumptions, we devised the following three-step algorithm.

1. To estimate the number of lights N and their respective azimuth coordinates ϕ_i, $i=1\ldots N$, we analyzed the intensity variation along the silhouette of the object. We assumed that the surface normals of the object at the silhouette lie in the image plane [13]. Using the silhouette normal assumption and the nominal diffuse lighting equation, we could accurately predict the azimuth coordinate ϕ of the individual lights. The number of lights and their relative intensities were estimated in an iterative fashion.
2. We used the globally convex assumption to estimate the zenith angles θ_i, $i=1\ldots N$ and relative intensities I_i. For each light detected in the first step, we swept the image from the silhouette to the interior along the azimuth direction, looking for maxima in the shading. The corresponding shape normal at the maxima $\vec{n}_i$ was indicative of the direction of the light and thus the zenith angle θ_i. To robustly handle local non convexities and back lighting, we detected and used shadows. Following Khan et al. [38], we differentiated the relatively high frequency variations of the luminance due to albedo (texture) from the low-frequency variations of luminance due to shading by using bilateral filtering.
3. By analyzing the median intensity in the shadow areas, we estimated the ambient light intensity.

Each of these steps is explained in detail in the following sections. However, we start by defining the coordinate system used. As depicted in Fig. 2d, the image plane was assumed to be aligned with the y–z plane, whereas the x-axis pointed out of the image plane. The origin lay at the center of the image. We also set a polar coordinate system such that the equator was aligned with the image plane and the axis was aligned with x-axis. Thus, the direction of a light was uniquely identified by the azimuth angle ϕ and the zenith angle θ.

4.2. Estimating azimuth angles

We assumed that the normals at the silhouette lay in the image plane. We further assumed that there were N discrete lights, each being either a directional light or a far-away point light (we estimate N below). Thus each light was uniquely characterized by its unknown luminance L_j and unknown unit direction $\boldsymbol{\omega}_j$, $j=1\ldots N$. To analyze the intensity variation of the silhouette pixels, we assumed a nominal Lambertian surface. Consider all pixels $\{\mathbf{p}_i\}$ that belong to the silhouette. Let $\mathbf{n}_i$ be the normal and L_i^v be the known luminance of the object at point $\mathbf{p}_i$:

$$L_i^v = \sum_{j=1}^{N} \Omega_{ij} L_j$$

$$\Omega_{ij} = \Omega(\mathbf{n}_i, \boldsymbol{\omega}_j) = \begin{cases} 0 & \text{if } \mathbf{n}_i \cdot \boldsymbol{\omega}_j < 0 \\ K_i^d \mathbf{n}_i \cdot \boldsymbol{\omega}_j & \text{if } \mathbf{n}_i \cdot \boldsymbol{\omega}_j \geq 0 \end{cases} \tag{1}$$

where K_i^d is the unknown diffuse reflectivity or albedo of pixel i. We encoded the normals, which were in the y–z plane, as polar coordinates $\phi_i^n \rightarrow \mathbf{n}_i = [0, \sin(\phi_i^n), \cos(\phi_i^n)]^T, 0 \leq \phi_i^n \leq 2\pi$.

To estimate the lights' azimuth angles ϕ_j^l, we used a k-means clustering algorithm. In traditional k-means clustering algorithms, each data point belongs to a certain cluster and affects the centroid of only that cluster. Unfortunately, a silhouette pixel may be illuminated by more than one light. Thus, we could not partition the pixels into exclusive clusters. Instead, we devised a partial voting scheme based on the Ω function to form 'fuzzy' clusters and to simultaneously compute the corresponding centroids as the lighting directions, as outlined in Algorithm 1.

Algorithm 1. Contour Voting—N lights

Require $\mathbf{L}^v \equiv \{L_i^v\}$ {discrete luminances}
Require $\mathbf{n} \equiv \{\mathbf{n}_i\}$ {silhouette normals}
Require $\boldsymbol{\phi}^n \equiv \{\phi_i^n\}$ {azimuth coordinates of the normals}
1: sort($\mathbf{L}^v, \mathbf{n}, \boldsymbol{\phi}^n$) {sort by decreasing luminances}
2: $\boldsymbol{\phi}^l \equiv \{\phi_j^l\} \mid j \in [1\ldots N]$ {azimuth coordinates of the lights}
3: seed($\boldsymbol{\phi}^l$)
4: $\boldsymbol{\alpha}^{\oplus} \equiv \{\alpha_j^{\oplus}\} \mid j \in [1\cdots N]$ {aggregate of weights per light}
5: $\boldsymbol{\alpha}^{\oplus} \leftarrow \mathbf{0}$
6: **Repeat**
7: **for all** $L_i^v \in \mathbf{L}^v$ **do**
8: $\boldsymbol{\omega}_j \leftarrow [0, \sin(\phi_j^l), \cos(\phi_j^l)]^T$ {current direction}
9: $\Omega_i^{\oplus} \leftarrow \sum_j \Omega(\mathbf{n}_i, \boldsymbol{\omega}_j)$ {total weight}
10: **for all** $j \in [1\ldots N]$ **do**
11: $\boldsymbol{\omega}_j \leftarrow [0, \sin(\phi_j^l), \cos(\phi_j^l)]^T$ {current direction}
12: $\alpha_{ij} \leftarrow L_i^v \Omega(\mathbf{n}_i, \boldsymbol{\omega}_j) / \Omega_i^{\oplus}$ {weight of normal i}
13: $\phi_j^l \leftarrow \alpha_j^{\oplus} \phi_j^l + \alpha_{ij} \phi_i^n$ {update direction}
14: $\alpha_j^{\oplus} \leftarrow \alpha_j^{\oplus} + \alpha_{ij}$
15: $\phi_j^l \leftarrow \phi_j^l / \alpha_j^{\oplus}$
16: **end for**
17: **end for**
18: **until** convergence($\boldsymbol{\phi}^l$)
19: **return** $\boldsymbol{\phi}^l$

We went through the list of pixels sorted by luminance (line 7) to perform the normal voting. Notice that each silhouette normal

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

ϕ_i^n votes for all the N light clusters (lines 10 to 16), according to their luminances L_i^v. However, each normal only partially votes for each light cluster, according to the Ω function (line 12). For that, the individual Ω function with respect to each light direction Ω_{ij} was normalized with the aggregate of the Ω functions $\Omega_i^{\oplus} = \sum_j \Omega(\mathbf{n}_i, \boldsymbol{w}_j)$.

We repeated the voting process (lines 7 to 17) until we converged on the light azimuth angles $\boldsymbol{\phi}^l$ (lines 6 and 18). The choice of the initial guess (line 3) for the azimuth angles was important to ensure a speedy and effective convergence. We assigned the azimuth of the brightest pixel's normal ϕ_1^n to the first light ϕ_1^l. For the successive lights, we set the azimuth angles to $\phi_1^l + 2\pi(j-1)/N$.

For the estimation of the number of lights N, our approach subsequently increased the number of lights $N = 1..i$ until either the error was below a given tolerance or the added light source did not improve the result. In practice, we found that the number of iterations was usually below $N = 4$. This was due to the quantization associated with the image's finite bit-depth. As the number of opposing lights increased, the variation in the shading over the surface decreased and became rather constant.

Although the proposed voting method has built-in resistance to local variations in albedo because of its search of global tendencies, ultimately, the results will be biased if the points in the contour form large clusters with very different luminance values, as the first image of Fig. 3a demonstrates.

It is possible to reduce this bias with a second pass, as follows. Once we have a set of N centroids (light directions), we went through all the voting pixels assigned to each k-group, corresponding to a light direction. We then checked that the dot product of the normal and the estimated light direction yielded a luminance value equal to the original luminance of the pixel, fractioned by its Ω function. If not, we forced the fractional albedo of the pixel to be coherent with the fractional luminance of the brightest pixel in the group. Then we repeated the contour voting algorithm. This correction in the albedo values usually produced small shifts (10–20°) in the directions in the case of extreme albedo variations (Fig. 3a).

As in other previous approaches based on contour analysis [39,40,14], the first step will fail if the light is situated around the x-axis; i.e., $\theta \approx \pi/2$. In this case there is no variation in luminances due to shading. This would result in erroneous estimation of the azimuth angles. However, the final direction of the light would be estimated accurately in the second step when we analyze the shading in the interior.

Finally, we corrected the potential bias along the direction stemming from the geometry of the silhouette. As depicted in Fig. 3b, a significant number of silhouette normals were parallel to the y-axis, biasing the resultant light towards that direction. We corrected this by eliminating multiple normals. We chose a set of discrete normal directions $\overline{\phi}_i^n$ and distributed all the silhouette normals into bins. Then, we computed the average luminance for each bin $\overline{L}_i$ and used this set of silhouette normals and luminances instead.

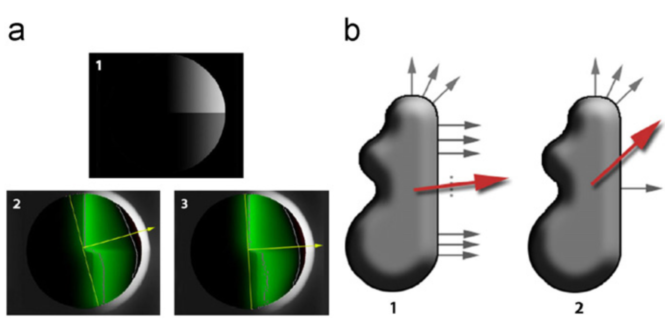

Fig. 3. (a^1) Sphere with a change in the albedo, (a^2) initial biased estimation because of a higher albedo, (a^3) corrected light direction estimate, (b^1) an estimate incorrectly biased because of the geometry of the silhouette, and (b^2) the correct result after eliminating multiple normals.

4.3. Estimating Zenith angles and intensities

To estimate zenith angles $\{\theta_j\}$ accurately, we disambiguated the luminance variations due to shading from the variations due to texture, which are relatively high in frequency. We used bilateral filtering to remove high frequencies while keeping lower frequency content, which is typically attributed to shading [38].

Then, for each light detected in the previous step, marching in the light's direction $\boldsymbol{\omega}_j = \boldsymbol{\omega}(\phi_j^l)$ from the silhouette to the interior, we analyzed the luminances. Because the pixels were lit by multiple lights, this directional derivative of the luminance $\boldsymbol{\omega}_j \cdot \nabla L^v$ was the main indicator of the shading from a particular light j aligned to its direction. There are two cases of luminance variations in the interior.

Case 1: If the directional derivative $\boldsymbol{\omega}_j \cdot \nabla L^v$ is positive at the silhouette, the light is directed towards the camera from the image ($\theta \geq 0$). In this case, the luminances continue to increase as we march along the direction of the light to reach the first local maximum. We denote this point as $\mathbf{p}_j^{hi}$. At this point, the surface normal points in the direction of the light; i.e., $\theta_j = \theta^n(\mathbf{p}_j^{hi})$. We ignore all the pixels thereafter because the geometry might be self-occluding or under the influence of another light.

Case 2: At the silhouette, if the directional derivative is negative, this is an indication of backlighting ($\theta < 0$). The luminances will successively decrease as we march along the light direction to reach a singularity. This point is the first self-shadow point $\mathbf{p}_j^{lo}$ and is marked by either a change of sign in the gradient of the directional derivative $\boldsymbol{\omega}_j \cdot \nabla L^v$ or a zero value of its luminance L^v. A change of sign will be produced when the contribution to the luminance value at that point by a second light is greater than the contribution of L^v. At this point, the surface normal is perpendicular to the light direction; i.e., $\theta_j - \theta^n(\mathbf{p}_j^{lo}) = \pi/2, \theta_j < 0$.

To estimate the normal at each point, we could not rely on shape-from-shading because of the overlapping of multiple lights. It was not possible to know a priori which combination of light sources was contributing to a certain point. Good solutions for estimating a valid normal at points $\mathbf{p}_j^{hi}$ or $\mathbf{p}_j^{lo}$ in arbitrary images do not exist [5].

Furthermore, this was complicated if two given points on the surface of the object were lit by a different and unknown number of light sources. Wang et al. [24] developed a technique to determine the number of lights, but they could do this thanks to accurate knowledge of 3D depth and normals. Instead, we reverted once more to our global convexity assumption and fit an ellipse along the scanline: one of the axes is given by the intersection of such a scanline and the silhouette; the other axis will approximate the object convexity and is a user parameter. By default, both axes are equal (in fact, defining a circumference). The surface normal was subsequently assumed to be the normal of the ellipse at the point under consideration.

We could start marching along the light direction from the brightest silhouette point that corresponds to the light. However, in order to minimize the influence of albedo variations, we scanned the light direction from multiple silhouette points. One way to realize this scheme was to rotate the image such that the light direction $\boldsymbol{\omega}(\phi_j^l)$ was aligned with the y-axis and the light on

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

ARTICLE IN PRESS

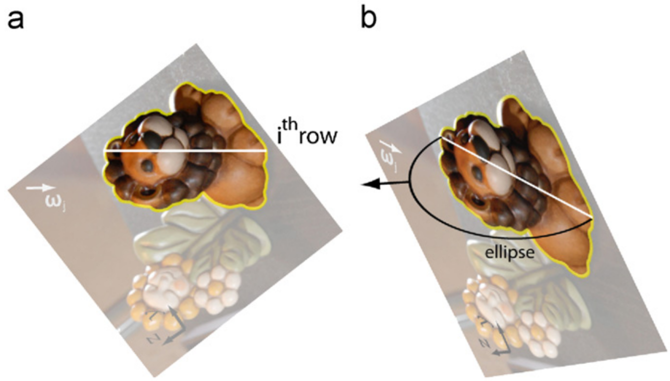

Fig. 4. Estimating zenith angle: (a) scanning in light direction for highlight or shadow and (b) ellipsoidal geometry.

the left, see Fig.4. Then, we simply scanned each raster line i, starting from the silhouette boundary on the left and moving into the interior. We detected the set of points $\{\mathbf{p}_{ij}^{hi}\}$ or $\{\mathbf{p}_{ij}^{lo}\}$ corresponding to the zenith angles $\{\theta_{ij}\}$ and the luminances L_{ij}^v. Thus, for the light j, the resultant zenith angle was the weighted sum:

$$\theta_j = \frac{\sum_i L_{ij}^v \theta_{ij}}{\sum_i L_{ij}^v} \tag{2}$$

By using two objects in the image as light probes and repeating the process for the second one, we could approximate the position of a point light source; i.e., a source that was not infinitely far away. Given that the directions computed for both objects could not be absolutely precise (we assumed directional lights), there was no intersection point. We simply placed the light source halfway between the points d_1 and d_2, defining its minimum distance.

Once we had estimates of the light directions, estimating the relative intensities was fairly straightforward. For each light j, we computed the total sum of the luminances normalized by the Ω function of the light over all the pixels i of the contour of the object. The intensity of the light I_j was proportional to

$$I_j \propto \sum_i L_j^v / \Omega(\mathbf{n}_i, \boldsymbol{\omega}_j). \tag{3}$$

Any potentially remaining light sources were treated by our algorithm as ambient illumination, which we will explain next.

4.4. Ambient illumination

The shading contribution of the ambient light was assumed to be constant for all pixels, and we could therefore estimate its intensity by analyzing pixels in the shadow regions. We had already detected the shadow lines in the previous step. The region bounded by these shadow lines was determined to be a shadow region. We averaged the set of samples along these boundaries. This ambient intensity estimate was also relative to the previously detected lights.

5. Results

Once we tested the accuracy of our method with real (controlled) light configurations, we further provided a visual validation of our method by using the lights detected in an image for automatic insertion and relighting of synthetic objects. Finally, we show a novel technique of image compositing based on our light detection method.

5.1. Error analysis

We have tested our algorithm on several images with controlled (known) light configurations to measure the errors in our light detection method. The images included varied configurations (see Fig. 5): Apple 1, Apple 2 and Apple 3 show a relatively simple geometry under very different lighting schemes (with one or two light sources, plus ambient light). The Guitar and Quilt images show much more complex scenes lit by three and two light sources, respectively. The light directions returned by our algorithm showed errors usually below 20° for the more restrictive azimuth angle ϕ, which is below the 30–35° limit set by our psychophysical findings. Even for the zenith angle θ, only the second light in the Quilt scene returned a larger error because of the bouncing of that light off the surface on the left. Table 1 shows all the data for the input images shown in Fig. 5. For each light source present in the scene, we show the real measured locations of the light sources, the results output by our algorithm and the corresponding absolute error. The number of directions was acquired automatically. The light probe used in the first three images was the apple; for the other two, we used the head of the guitar player and the Scottish quilt.

We can select multiple objects (or convex parts of objects) in a single image as light probes, as shown in Fig. 6. In these cases, the analysis returns coherent results for global light sources. Local sources may spatially vary in the image. In both cases (Apollo's arm and the body of Vulcan's assistant), the main light shows almost the same direction. This figure also shows the applicability of our method to 2D paintings. In this case we can observe how the artist intended (and was able) to have both characters under consistent lighting.

5.2. Visual validation

We further tested our algorithm on uncontrolled images, depicting scenes with unknown illuminations and varying degrees of diffuse-directional lighting ratios. Given that we obviously cannot provide error measures in those cases, we provide visual validation of the results by rendering a synthetic object with the lighting scheme returned by our algorithm. Fig. 7, left, shows the original image and an untextured version of the 3D

Fig. 5. Input images for the error analysis of Table 1. From left to right: Apple 1, Apple 2 and Apple 3, guitar and quilt.

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

objects to be rendered. The image on the right shows the results of illuminating the 3D objects with the output returned by our algorithm. The chosen light probe was one of the mushrooms. Fig. 8 shows additional examples of uncontrolled input images with synthetic objects rendered into them; the head of the doll and the whole human figure were used as light probes, respectively. Note how our system is robust enough even if the light probe is composed of multiple objects with very different BRDFs (such as the skin, glasses and hair in the doll image). The shadows cast onto the original images were generated by shadow mapping and synthetic planes manually set at approximately the right locations when placing the synthetic objects.

5.3. Image compositing

Finally, we applied the illumination information obtained by our method to a well-known problem in computer graphics: compositing two images with different illumination environments into a single image with coherent illumination. In image compositing, color and intensity can be adjusted with relatively straightforward techniques including Poisson-based approaches [41] and color transfer algorithms [42]. Although such algorithms go a long way toward matching the color schemes, they do not match the illumination direction on objects. Thus, if strong localized lighting exists in either the source or the target images, the result will look out of place.

For compositing we used the following approach: first, we analyzed the background image with our light detection method. Second, we extracted a coarse 3D shape of the image to be inserted. Third, we relit this shape using the lights' directions and intensities from the first step and pasted it in the final image.

We first needed to produce a plausible depth map of every object in the scene to be relit. This can be achieved in a number of ways [43,6], but we chose to follow a simple method [38] based on the interpretation of luminance as depth values [38]. This approach has been successfully used before in the context of image-based material editing [38] or light transport editing [44]. A bilateral filter [45] was applied to the result to remove high-frequency details. The obtained depth values $D(x,y)$ represented the camera-facing half of the object. For image relighting, we additionally needed an approximation of the far side of the object, which aids in the casting of shadows and the computation of diffuse interreflections. As our input did not allow us to infer this geometry with any decent accuracy, we reconstructed this backfacing geometry simply by mirror-copying the front half of the recovered geometry, in accordance with our global convexity assumption. Again, the obvious inaccuracies of this approach are masked by the limitations of our visual perception, as our final results show. To prepare our recovered geometry for relighting, we finally computed a normalized surface normal $n(x,y)$ for each pixel belonging to the object from the gradient field $\nabla z(x,y)$.

Once the 3D shape is known, several rendering approaches are available, and there are no limitations on the complexity of the BRDF employed. For demonstration purposes, we used a combination of Lambert's and Phong's models to represent the surface reflectance [46]. The new texture of the object was generated from the original image using the original hue and saturation channels and the high-frequency component of the original luminance channel (extracted by means of a bilateral filter [38]). Figs. 9 and 10 show examples of the aforementioned relighting technique, which was used in combination with light detection to obtain the composition of the flute in Fig. 11. As input for the relighting phase and because of the white balance/albedo ambiguity in the lightprobe, the user has to set

Table 1
Real measured light directions (R), value returned by our algorithm (A) and absolute error (E) for the zenith θ and azimuth ϕ angles in the scenes depicted in Fig. 5.

	Light 1		**Light 2**		**Light 3**	
	ϕ	θ	ϕ	θ	ϕ	θ
Apple 1						
R	−15.00	40.00	165.00	−40.00	–	–
A	5.71	35.31	162.25	−64.03	–	–
E	**20.71**	**4.69**	**2.75**	**24.03**	–	–
Apple 2						
R	90.00	−70.00	–	–	–	–
A	94.54	−65.70	–	–	–	–
E	**4.54**	**4.3**	–	–	–	–
Apple 3						
R	180.00	0.00	0.00	0.00	–	–
A	168.50	14.48	0.0	11.31	–	–
E	**12.50**	**14.48**	**0.00**	**11.31**	–	–
Guitar						
R	180.00	10.00	30.00	−45.00	260.00	45.00
A	185.71	29.66	25.64	−49.19	272.29	41.48
E	**5.71**	**19.66**	**4.36**	**4.19**	**12.29**	**3.16**
Quilt						
R	10.00	−35.00	120.00	−10.00	–	–
A	24.70	−51.79	162.25	4.74	–	–
E	**14.70**	**16.79**	**42.25**	**14.74**	–	–

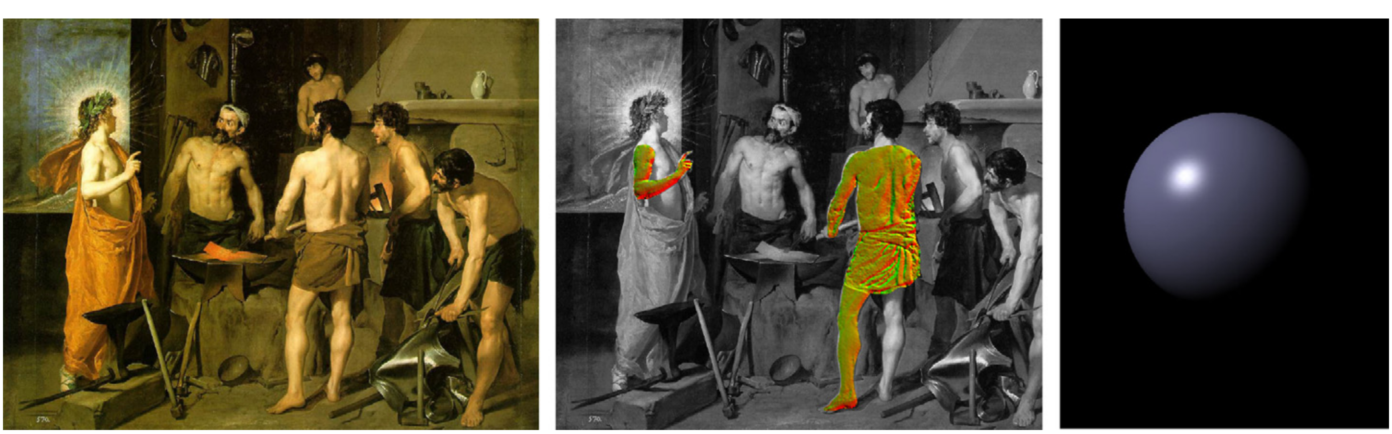

Fig. 6. Left: input image, La fragua de Vulcano by Diego de Velazquez (1630), oil on canvas. Middle: areas used as light probes showing the computed horizontal (red) and vertical (green) gradients. Note how the user can select parts of an object, avoiding, for instance, the black albedo of the hair on the head or the shadows in the right leg. Right: a synthetic OpenGL render with the light source detected for the arm. The light direction was estimated as $(\phi,\theta) = (139.97,33.04)$ for the arm and $(\phi,\theta) = (136.17,39.10)$ for the body. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

a base luminance level and a color per light source. The directions and relative intensities are provided by our method. In our experiments we found that this kind of input was feasible for an unskilled user if the tuning was done interactively once the object was inserted with all the lights set as white by default.

Fig. 7. Rendering synthetic objects into the images. Left, top: original input image (light probe highlighted). Left, bottom: 3D models lit according to the output of our light detection algorithm. Right: final result with the 3D models textured and inserted into the image.

6. Discussion and future work

We have presented a novel light detection algorithm for single images that only requires the silhouette of any object in the image as additional user input. Our method yields a result in less than 4 s using a 512×512 version of the original image. Although it works on lower resolution images, higher-resolution images have a smaller effect on the accuracy of the technique. It may seem that the average error of our method is too high in comparison with previous works in the field; however, compared with those works, we are not limited to detecting just one light source, and no knowledge of the actual 3D geometry is required. Moreover, our psychophysical study confirmed that our results are below a threshold where illumination inconsistencies tend to go unnoticed by human vision.

We have shown good results both with controlled lighting environments (where the light positions were measured and thus numerical data could be compared) and uncontrolled settings (with free images downloaded from the internet and with synthetic objects rendered with the results of our algorithm). Furthermore, we have introduced a novel image compositing

Fig. 8. Additional examples of synthetic objects rendered into images using the results of our algorithm. Left: synthetic teapot. Right: synthetic cone.

Fig. 9. Two new images relit with our method. Inset: original image.

Fig. 10. A more dramatic lighting change. Left, original image. Right, the altered version, resembling moonlight as it would possibly be shot by a cinematographer.

Fig. 11. Demonstration of our compositing method. The crumbled papers were chosen as light probe. From left to right: original image. Image of a flute to be inserted. Final composition after light detection and relighting.

method based on our light detection method. Our algorithm could help photographers mimic a given lighting scheme inspired by any other shot for which a reduced set of light directions (namely, the typical three-light setup made up of key, fill and rim lights) is preferable.

It could be argued that because humans are not particularly good at detecting light sources, simpler algorithms that approximate light sources could be employed instead. For instance, in the context of rendering synthetic objects into existing images, one of the most popular recent approaches is to build an environment map from the image. While this approach would provide reasonable results in certain cases (as shown in [38]), it would fail if the main light sources were actually outside the image. One such example would be the Guitar image in Fig. 5. If we were to render an object into the image, it would appear unrealistically dark. Fig. 12 shows a sphere rendered with the actual measured lights for that scene compared with the results from rendering with an environment map and using the lights detected by our algorithm.

Several existing applications could benefit from our system, specifically those based on combining pictures from an existing stack to create novel images. These kinds of applications are gaining popularity because of, among other factors, the existence of huge databases and their accessibility through the internet. Some examples include Photo Clip Art [7], Interactive Digital Photomontage [8] and Photo Tourism [9].

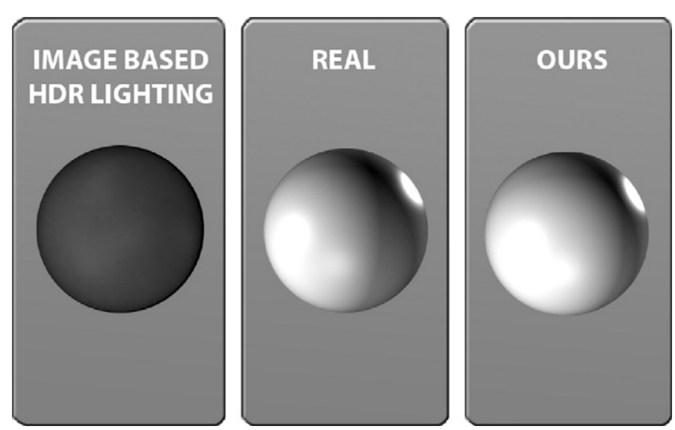

Fig. 12. Spheres rendered with information from the Guitar image in Fig. 5. Left: using the image as an environment map. Middle: using the real measured data. Right: using the results of our algorithm. Our algorithm provides a much better solution if the light sources are not present in the original image.

We assumed global convexity for the chosen de facto light probes in the images. Although this assumption is true for most objects, the algorithm will return wrong values if a concave object is chosen instead. Our algorithm will also fail in the presence of purely

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

reflective or transparent (refractive) objects chosen as light probes, which break our assumption about shading. In these cases, an approach similar to [22] may be more suitable, although previous knowledge about the geometry of the objects in the image would be needed. As future work, we would like to address these cases.

Additionally, the novel compositing method introduced in a previous section has three aspects that need further research. First the recovered 3D shape is obtained by means of a simple shape derived from the shading approach, which might produce wrong and unexpected results with certain light configurations. Given the plausible results we obtained with such a simple method, we intend to test more sophisticated 3D shape recovery algorithms [47,48]. Second, regarding the recovered texture of the object to be relit, our approach is valid for images in which the original hue and saturation values are available for most pixels. This assumption works in our examples where shadows are not harsh or cover a small portion of the image (frontal flashlight) or when high dynamic range information is available (hue and saturation values are captured even for pixels in low luminance areas). Hence, for a broader range of scenarios, we plan to research approaches like learning-based filtering and belief propagation [49] to obtain a more accurate separation between the reflectance and the shading of the object before the relighting process. Finally, we intend to validate our composition results by means of additional psychophysical studies.

Acknowledgements

This research was funded by a Marie Curie grant from the Seventh Framework Programme (Grant agreement no.: 251415), a generous gift from Adobe Systems Inc., the Spanish Ministry of Science and Technology (TIN2010-21543) and the Gobierno de Aragón (Projects OTRI 2009/0411 and CTPP05/09).

References

[1] Debevec P. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: SIGGRAPH '98: proceedings of the 25th annual conference on computer graphics and interactive techniques. New York, NY, USA: ACM; 1998. p. 189–98. doi:〈http://doi.acm.org/10.1145/280814.280864〉.
[2] Jacobs K, Nielsen AH, Vesterbaek J, Loscos C. Coherent radiance capture of scenes under changing illumination conditions for relighting applications. The Visual Computer 2010;26(3):171–85. doi:〈http://dx.doi.org/10.1007/s00371-009-0360-2〉.
[3] Gibson S, Howard T, Hubbold R. Flexible image-based photometric reconstruction using virtual light sources. Computer Graphics Forum 2001;19(3):C203–14.
[4] Madsen CB, Nielsen M. Towards probe-less augmented reality—a position paper. In: GRAPP, 2008. p. 255–61.
[5] Zhang R, Tsai P, Cryer J, Shah M. Shape from shading: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 1999;28(8):690–706.
[6] Oh BM, Chen M, Dorsey J, Durand F. Image-based modeling and photo editing. In: SIGGRAPH '01: proceedings of the 28th annual conference on computer graphics and interactive techniques, 2001. p. 433–42.
[7] Lalonde J-F, Hoiem D, Efros AA, Rother C, Winn J, Criminisi A. Photo clip art. ACM Transactions on Graphics (SIGGRAPH 2007) 2007;26(3).
[8] Agarwala A, Dontcheva M, Agrawala M, Drucker S, Colburn A, Curless B, et al. Interactive digital photomontage. ACM Transactions on Graphics 2004;23(3):294–302.
[9] Snavely N, Seitz SM, Szeliski R. Photo tourism: exploring photo collections in 3d. ACM Transactions on Graphics 2006;25(3):835–46.
[10] Pentland A. Finding the illuminant direction. Journal of the Optical Society of America A 1982;72(4):448–55.
[11] Brooks M, Horn B. Shape and source from shading. In: Proceedings of the international joint conference on artificial intelligence, 1985. p. 932–6.
[12] Lee C, Rosenfeld A. Improved methods of estimated shape from shading using the light source coordinate system. In: Horn B, Brooks M, editors. Shape from shading. MIT Press; 1989. p. 323–569.
[13] Horn B. Robot vision. McGraw-Hill; 1986.
[14] Nillius P, Eklundh J-O. In: CVPR, 2001. p. I:1076–83 〈citeseer.ist.psu.edu/nillius01automatic.html〉.
[15] Koenderink JJ, Pont SC. Irradiation direction from texture. Journal of the Optical Society of America 2003;20(10):1875–82.
[16] Varma M, Zisserman A. Estimating illumination direction from textured images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Washington, DC, vol. 1, 2004. p. 179–86.
[17] Hougen D, Ahuja N. Estimation of the light source distribution and its use in integrated shape recovery from stereo shading. In: ICCV, 1993. p. 29–34.
[18] Zhang Y, Yang Y-H. Multiple illuminant direction detection with application to image synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001;23(8):915–20. doi:〈http://dx.doi.org/10.1109/34.946995〉.
[19] Powell M, Sarkar S, Goldgof D. A simple strategy for calibrating the geometry of light sources. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001;23(9):1022–7. doi:〈http://doi.ieeecomputersociety.org/10.1109/34.955114〉.
[20] Lagger P, Fua P. Using specularities to recover multiple light sources in the presence of texture. In: ICPR '06: proceedings of the 18th international conference on pattern recognition. Washington, DC, USA: IEEE Computer Society; 2006. p. 587–90. doi:〈http://dx.doi.org/10.1109/ICPR.2006.1156〉.
[21] Zhou W, Kambhamettu C. Estimation of illuminant direction and intensity of multiple light sources. In: ECCV'02: proceedings of the seventh European conference on computer vision—part IV London, UK: Springer-Verlag; 2002. p. 206–20.
[22] Nishino K, Nayar SK. Eyes for relighting. ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH) 2004;23(3):704–11.
[23] Marschner SR, Greenberg DP. Inverse lighting for photography. In: Fifth IST/SID color imaging conference, 1997. p. 262–5.
[24] Wang Y, Samaras D. Estimation of multiple illuminants from a single image of arbitrary known geometry. In: ECCV02, vol. 3, 2002. p. 272–88.
[25] Sato I, Sato Y, Ikeuchi K. Illumination distribution from brightness in shadows: adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In: International conference on computer vision (ICCV), vol. 2, 1999. p. 875–82.
[26] Dror RO, Leung TK, Adelson EH, Willsky AS. Statistics of real-world illumination. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Kauai, Hawaii, 2001.
[27] Pouli T, Cunningham D, Reinhard E. EUROGRAPHICS 2010 STAR: image statistics and their applications in computer graphics, 2010.
[28] Dror RO, Willsky AS, Adelson EH. Statistical characterization of real-world illumination. Journal of Vision 2004;4:821–37.
[29] Mury AA, Pont SC, Koenderink JJ. Light field constancy within natural scenes. Applied Optics 2007;46(29):7308–16.
[30] Langer MS, Bülthoff HH. A prior for global convexity in local shape-from-shading. Perception 2001;30:403–10.
[31] Ostrovsky Y, Cavanagh P, Sinha P. Perceiving illumination inconsistencies in scenes. Perception 2005;34:1301–14.
[32] Enns J, Rensik R. Influence of scene-based properties on visual search. Science 1990;247:721–3.
[33] Kleffner D, Ramachandran V. On the perception of shape from shading. Perception and Psychophysics 1992;52:18–36.
[34] Lopez-Moreno J, Sangorrin F, Latorre P, Gutierrez D. Where are the lights? Measuring the accuracy of human vision. In: CEIG '09: Congreso Español de Informática Gráfica, 2009. p. 145–52.
[35] Lopez-Moreno J, Sundstedt V, Sangorrin F, Gutierrez D. Measuring the perception of light inconsistencies. In: APGV '10: proceedings of the seventh symposium on applied perception in graphics and visualization. ACM; 2010. p. 25–32.
[36] Koenderink JJ, van Doorn AJ, Pont SC. Light direction from shad(ow)ed random Gaussian surfaces. Perception 2004;33(12):1405–20.
[37] Wang J, Agrawala M, Cohen M. Soft scissors: an interactive tool for realtime high quality matting. ACM Transactions on Graphics 2007;26(3).
[38] Khan EA, Reinhard E, Fleming R, Bülthoff H. Image-based material editing. ACM Transactions on Graphics (SIGGRAPH 2006) 2006;25(3):654–63.
[39] Yang Y, Yuille A. Sources from shading. In: Computer vision and pattern recognition, 1991. p. 534–9.
[40] Vega E, Yang Y-H. Default shape theory: with the application to the computation of the direction of the light source. Journal of the Optical Society of America A 1994;60:285–99.
[41] Jia J, Sun J, Tang C-K, Shum H-Y. Drag-and-drop pasting. In: SIGGRAPH '06: ACM SIGGRAPH 2006 papers. New York, NY, USA: ACM; 2006. p. 631–7. doi:〈http://doi.acm.org/10.1145/1179352.1141934〉.
[42] Reinhard E, Ashikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Computer Graphics and Applications 2001;21(5):34–41.
[43] Igarashi T, Matsuoka S, Tanaka H. Teddy: a sketching interface for 3d freeform design. In: SIGGRAPH '99: proceedings of the 26th annual conference on computer graphics and interactive techniques, 1999. p. 409–16.
[44] Gutierrez D, Lopez-Moreno J, Fandos J, Seron F, Sanchez M, Reinhard E. Depicting procedural caustics in single images. ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia) 2008;27(5):1201–9.
[45] Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: Proceedings of the IEEE international conference on computer vision, 1998. p. 836–46.

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004

[46] Foley J, van Dam A, Feiner S, Hughes J. Computer graphics principles and practice. 2nd ed. Addison-Wesley; 1990.

[47] Samaras D, Metaxas D. Coupled lighting direction and shape estimation from single images. In: ICCV '99: proceedings of the international conference on computer vision, vol. 2. Washington, DC, USA: IEEE Computer Society; 1999. p. 868.

[48] Durou J-D, Falcone M, Sagona M. Numerical methods for shape-from-shading: a new survey with benchmarks. Computer Vision and Image Understanding 2008;109(1):22–43.

[49] Tappen MF, Freeman WT, Adelson EH. Recovering intrinsic images from a single image. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005;27(9):1459–72.

Please cite this article as: Lopez-Moreno J, et al. Compositing images through light source detection. Computers and Graphics (2010), doi:10.1016/j.cag.2010.08.004