Shadow-based Light Detection for HDR Environment Maps

Andrew Chalmers and Taehyun Rhee
Computational Media Innovation Centre (CMIC)
Victoria University of Wellington
Wellington, New Zealand
{andrew.chalmers, taehyun.rhee}@vuw.ac.nz

***Abstract*—High dynamic range (HDR) environment maps (EMs) are spherical textures containing HDR pixels used for illuminating virtual scenes with high realism. Detecting as few necessary pixels as possible within the EM is important for a variety of tasks, such as real-time rendering and EM database management. To address this, we propose a shadow-based algorithm for detecting the most dominant light sources within an EM. This algorithm takes into account the relative impact of all other light sources within the upper-hemisphere of the texture. This is achieved by decomposing an EM into superpixels, sorting the superpixels from brightest to least, and using $\ell0$-norm minimisation to keep only the necessary superpixels that maintains the shadow quality of the EM with respect to the just noticeable difference (JND) principle. We show that our method improves upon prior methods in detecting as few lights as possible while still preserving the shadow-casting properties of EMs.**

***Index Terms*—light detection, shadow, high dynamic range, environment map, lighting**

I. Introduction

Environment maps (EMs) are high dynamic range (HDR) 360° textures that are suitable for storing real-world lighting. Image-based lighting (IBL) [1] is a rendering process that uses EMs for lighting a virtual scene. This rendering process is expensive and is often used for offline applications such as films. However, real-time applications can adopt IBL with a few optimisations, such as light detection and pre-convolving EMs for efficient shading. Of interest to this paper is the light detection, where the high frequency pixels are converted into directional light sources for real-time shadow casting. Another application area is regarding databases of EMs, where EMs require automatic tagging of their high frequency lighting properties [2], [3]. A method for detecting bright pixels within EMs, based on how they cast shadows, is necessary to accurately represent EMs in real-time rendering and database management applications.

Previous methods are either not designed for HDR [4], [5] or do not take into each light's relative impact to one another within the hemisphere of the EM [6]. To overcome these limitations, we design an algorithm that detects an HDR EM's most dominant light sources by analysing the illumination properties of the EM. To do this, we use the shadow-casting properties of EMs since they are camera independent while still being able to capture the illumination effects of bright light sources. We also use a carefully chosen rendered scene (cylinder-plane) that takes into account the upper-hemisphere of the EM, as well as taking into account the relative impact that each light source has on one another in the EM. In doing so, this algorithm does not detect all light sources, but instead detects a concise set of only the brightest light sources. We use the *just noticeable difference* (JND) to perceptually inform our algorithm on which light source have a visually significant impact on the rendered shadow quality. The end result is an mask for the EM that indicates which pixels have the most significant impact on the shadows in the rendered scene.

This project is funded by MBIE Smart Ideas and in part TEC.

978-1-7281-8579-8/20/$31.00 ©2020 IEEE

II. Related Work

Previous work in this area has focused on LDR EMs [4], [7]. These methods had to work around the problem of differentiating between light sources and reflections within images, whereas other methods [8] do not differentiate between the two. This is due to the fact that the LDR image format does not contain reliable luminance values, and therefore the pixel intensity values for light sources and reflections are identical. In order to differentiate between the two, solutions are based on the textural properties of the image and supervised learning.

A related field is compression and sampling methods. Compression algorithms can reduce an EM into a small number of coefficients [9]–[11], but the brightest lights are not described efficiently, since the coefficients will attempt to describe both high and low frequency properties of the EM. An exception is that spherical harmonics can be used to obtain a single dominant light direction [12], [13], however the area (not just the direction) of the lights is a sought after property in determining the softness of the illumination effect. Sampling algorithms sample areas of interest in the EM [14]–[17], however, these sampling methods aim to oversample areas of interest rather than provide a clear solution of which pixels need to be regarded as the most dominant light sources.

We are interested in describing multiple lights within an EM, as well as describing other properties of a light, such as its area. Recent work [4], [18] has been able to obtain a mask that extracts segments of the EM which correspond to light sources. In this paper, we propose a new light detection algorithm suitable for HDR EMs, and detect only light sources which have a meaningful impact on the rendered scene. Furthermore, detecting a smaller number of more important light sources, suitable for various application areas such as EM database management and real-time rendering.

III. Dominant Light Detection

In this section we define what a dominant light is, as well as the algorithm that detects these dominant light sources.

A. Dominant Light Definition

An EM comprises of pixels each representing directional light sources. We define dominant light sources are pixels within the EM that contribute to the specular highlights and shadows. Both shadows and highlights are illumination properties that correspond to a light source, of which we opt to primarily use shadows in our light detection algorithm. This is motivated by the fact that shadows can be used to reliably reconstruct the high frequency lighting for varying eye position or materials [19], as opposed to highlights which are dependent on them [9]. Specifically, a dominant light in an EM is a local region of the lighting sphere $\Omega \subset \mathbb{S}^2$ with the property that the region has a substantial impact on high-frequency lighting effects in a scene, including highlights and cast shadows.

B. Dominant Light Detection Algorithm

The shadow information in the EM's corresponding rendered image is used as an indication of whether or not pixels in the EM represent a dominant light source which caused the shadows. We propose an iterative algorithm for determining whether or not pixels in a EM are dominant light sources.

For a given EM, we render a ground truth image:

$$I(\mathrm{r}) = \int_{\Omega} L_i(\omega_i) f(\omega_i) \left|\cos\theta\right| \, \mathrm{d}\omega_i, \tag{1}$$

where I is the rendered image given a EM r which computes the rendering equation. L_i is a light sampled from EM r. We use a top-down orthographic camera of a cylinder on a ground plane, as shown in Figure 1. The cylinder casts shadows in each direction, and the ground plane captures the shadow information. We use a cylinder because it is symmetrical and casts shadows evenly in each direction. The material property f is a diffuse Lambertian material for both the cylinder and ground plane.

From the ground truth rendering, we then extract the shadow information. To do this, we adopt a differential rendering approach [20] and compute the difference between the rendered image of a cylinder on the plane with a rendered image of only the plane. This in effect is subtracting the diffuse value of the scene from the image, producing shadow extracted image in Figure 2:

$$f(I) = I - I_d, \tag{2}$$

where I is the rendered image and I_d is the rendering of the Lambertian reflectance model with respect to the normal of the ground plane. The colour of the material is set to white. The result of this operation is a difference image which contains all the shadow detail (see Figure 2).

However, there are a lot of shadow details stored in $f(I)$ that we cannot perceive. Our goal is to minimise the complexity of a EM by detecting the smallest number of light sources required to reconstruct the ground truth rendered image. Therefore, we modify the difference image by removing shadow information which we cannot perceive. This has an effect on the optimisation such that we are not fitting the ground truth of all the shadow detail, but instead, we are fitting shadows that we can only perceive. This in effect will discriminate against non-dominant light sources which do not cast shadows in the perceived shadow region.

We use the just noticeable difference (JND) and Weber's Law [21] to determine which shadows are perceivable to the human eye:

$$\frac{\Delta I_d}{I_d} = k, \tag{3}$$

where I_d is the diffuse unoccluded intensity value, and ΔI_d is the change in intensity which is perceivable. It is shown that the ratio between the diffuse light and the perceived additional light is constant k. Therefore, we use this law to act as a threshold value to eliminate shadow regions which are not perceivable to the human visual system. For the human visual system, k is approximately in the range of $\frac{1}{12}$ to $\frac{1}{100}$ [22]. Experimentally we found that $\frac{1}{30}$ worked well. In practice, this is a threshold value on the difference image. Therefore, the JND threshold is k scaled by the diffuse intensity

$$\text{JND} = k \cdot I_d, \tag{4}$$

which we apply to $f(I)$

$$g(f(I)) = \begin{cases} 0, & f(I) \leq \text{JND} \\ f(I), & \text{otherwise.} \end{cases} \tag{5}$$

For brighter or dimmer surfaces, the JND threshold scales appropriately. See Figure 2 for an example of the rendered images and the visible difference of the shadow region.

Once the ground truth data is established, we then compute superpixels using SLIC [23] on the EM we aim to detect the lights from. We consider each superpixel as a potential dominant light source. From this set, we need to determine if the superpixel is a shadow-casting or shadow-altering light source. We use a variation of an l_0 optimisation to obtain the optimal set. We find the smallest set of superpixels such that we reconstruct the JND filtered difference image:

$$\begin{aligned} \min_{x} \quad & \|x\|_0 \\ \text{subject to} \quad & f(I(Ax)) = g(f(I(b))), \end{aligned} \tag{6}$$

where A is a matrix where each column is a EM for each superpixel, b is the rendered image by the unaltered input EM, and x is a binary activation function of the superpixels in A. l_0-optimisation is NP-hard as there is a large number of combinations in A to approximate b. However, to improve the iterative

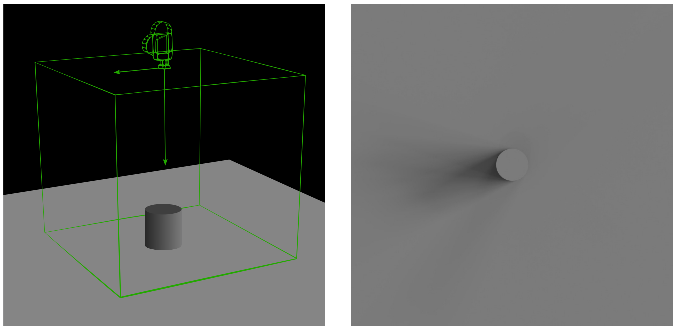


Fig. 1: On the left is a top-down orthographic camera of a cylinder on a plane. This setup is used to obtain a clear view of the shadow information cast in all directions on the horizontal plane, and on the right is an example rendering using an EM as a light source.

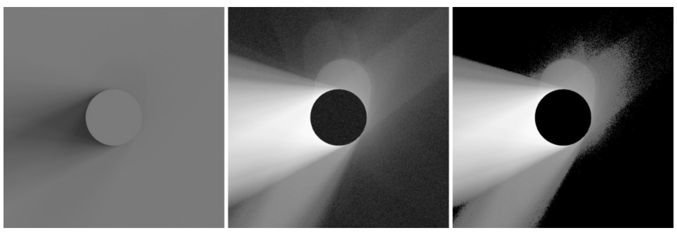


Fig. 2: From the left: rendered image, difference image showing shadow detail of the cylinder on the ground plane, and thresholding to remove shadows not perceivable to the human eye using Weber's Law. These show a zoomed in version of the data that was used.

algorithm, we sort the superpixels in descending order, from the highest intensity superpixel to the least. From this, we use a brute-force matching pursuit algorithm and enable each superpixel sequentially. We compute the difference image of the rendering at each step of the iteration, and compare it with the ground truth shadow difference image. The error is computed using STSIM-2 [24], [25], which is a metric that corresponds with the human visual system and has shown good results compared with other perceptual metrics [26]. If the similarity score improves on the previous iteration, the current superpixel is kept, if it does not improve, it is rejected. Because we are discriminating against shadow information that we cannot perceive using Weber's Law, we are in effect discriminating against light sources that casts shadows we cannot perceive with the human visual system. This has the desirable effect of omitting light sources in the EM that do not emit enough radiant intensity to cast perceivable shadows. There is a free parameter which can be added to the error to discriminate against superpixels which do not increase the error, but also do not improve the error by very much. In practice, we use a small value 0.0001.

Once we iterate through each superpixel, a set of superpixels are found as the dominant light sources (see Figure 3 for an example). These are the bare minimum superpixels required to produce a rendered image with shadow information which

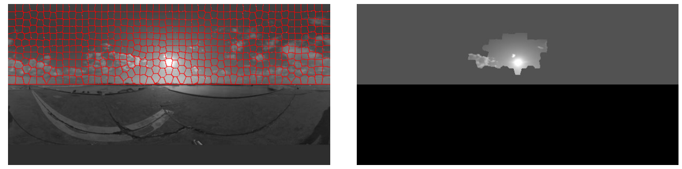


Fig. 3: Our light detection algorithm computes superpixels on the EM (left) and finds the set of superpixels which contribute as a dominant light source (right).

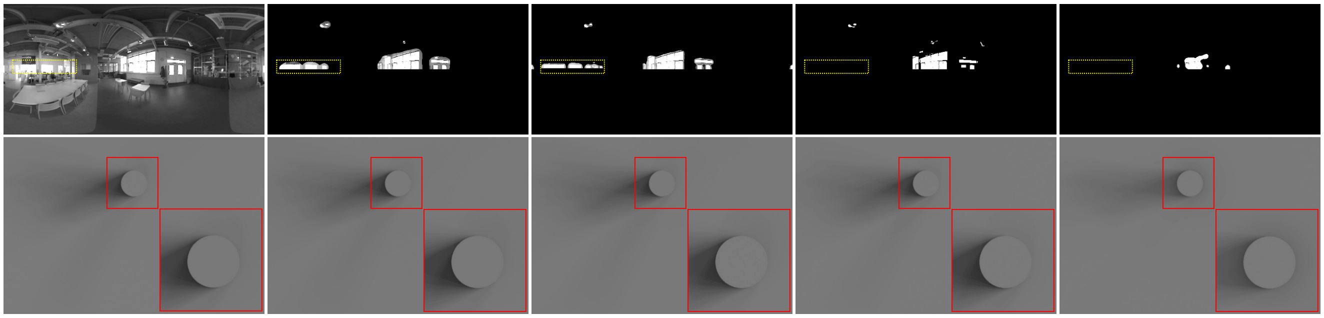
looks very similar to rendered image by the original EM. Every superpixel adjacent to another is combined into a single patch. Each patch area is considered a dominant light source. It is possible for multiple dominant light sources to share the same patch region. To avoid such cases, local maximas are computed on the light detected EM. Instead of assigning superpixels to adjacent superpixels, the superpixels are assigned to the nearest local maxima. To improve the local maxima detection, the noise is reduced using a Gaussian blur filter. Each feature for each light source having the azimuth position F_{lx} is set to the x position of the local maximas. See Section IV for the results of our dominant light detection algorithm.

IV. Evaluation

We compared our dominant light detection with that of Karsch et al. [4] (modified for HDR data), as well as to naive thresholding. Figure 4 illustrates how our method detects only the necessary pixels in order to obtain all the perceivable shadow detail compared to the other methods. We also conducted a user study where 50 participants were asked to compare the ground truth rendered image with the dominant light detected EMs of each method using a 7-point Likert item similarity scale. See Figure 5 for the results of our user study, which show that our method maintains high quality compared with the ground truth while also reducing the the number of active EM pixels. We can observe that thresholding also provides high quality results, but this is due to the method not removing enough unnecessary pixels (i.e. under-detecting). Our method successfully maintains high quality while also removing 95.6% pixels from an EM on average.

V. Conclusion

This paper presents a light detection algorithm designed for HDR EMs. The algorithm is designed to detect the most dominant light sources, to aid in the use of various applications such as EM database management and real-time rendering. With this consideration, we carefully designed a rendered scene that captures the shadows and relative impact of lighting on the upper hemisphere of the EM. A matching pursuit greedy algorithm is employed to detect the pixels in the EM that retains the shadow quality in the rendered scene. Each iteration in the matching pursuit algorithm requires a rendered scene, which takes time. We minimised this issue by using a greedy solver and low resolution renderings. Future work can look at optimised path tracing algorithms to speed up this up. The focus of this paper is detecting a mask of pixels. Future work



(a) Ground truth (b) Thresholding $T = 1.0$ (c) [Karsch14] (d) Our method (e) Thresholding $T = 15.0$

Fig. 4: Comparison of dominant light detection methods (top row) and their illumination effects in the rendered image (bottom row). (a) is the ground truth, (b) is naive thresholding ($T = 1.0$), (c) is Karsch's method, and (d) is our method. Notice that the rendered images look similar to the ground truth, yet our method does not detect unnecessary light sources as illustrated in the yellow dotted region. (e) is thresholding by increasing the threshold ($T = 15.0$) to remove the unnecessary light sources, yet this introduces artifacts in the rendered scene (shadow detail is lighter and softer) due to removing necessary light sources.

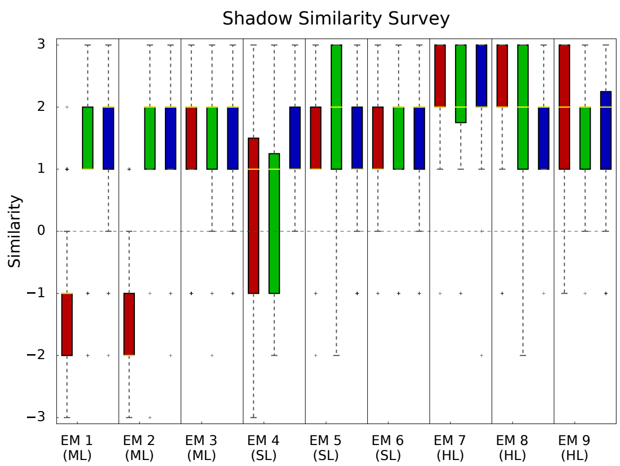


Fig. 5: User study results comparing [Karsch14] light detection (red bars), thresholding (green bars) and our method (blue bars). The reduced amount of dominant light pixels averaged across for each EM is 79.3%, 80.1% and 95.6% for each light detection method respectively. ML are the EMs with multiple lights, SL has one soft light, and HL has one hard light. Our approach maintains high quality while also removing the most pixels.

can also consider transforming the mask of pixels into abstract light sources (e.g., directional and area light sources) for real-time rendering. Our solution can directly be used for detecting the lights in the *Illumination Browser* [2] for EM database management. The shadow-based scheme can be considered for mobile mixed reality applications [27] where shadows are the only visible information to estimate the lighting properties.

Acknowledgement

This project was funded by the Smart Ideas Endeavour Fund from MBIE and in part by the Entrepreneurial University Programme from TEC in New Zealand.

References

[1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, 2008, pp. 1–10.

[2] A. Chalmers, J. Zhao, D. Medeiros, and T. Rhee, "Reconstructing reflection maps using a stacked-cnn for mixed reality rendering," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[3] C. Bloch, C. Bauer, V. Heisterberge, G. Reddick, and chris Huf, "Smart-ibl," http://www.hdrlabs.com/sibl/, 2008, accessed: 30-06-2020.

[4] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, H. Jin, R. Fonte, M. Sittig, and D. Forsyth, "Automatic scene inference for 3d object compositing," *ACM Trans. Graph.*, 2014.

[5] T. Rhee, L. Petikam, B. Allen, and A. Chalmers, "Mr360: Mixed reality rendering for 360 panoramic videos," *IEEE Transactions on Visualization & Computer Graphics*, no. 4, pp. 1379–1388, 2017.

[6] M. Laskowski, "Detection of light sources in digital photographs," in *11th Central European Seminar on Computer Graphics*, 2007.

[7] J. Xiao, K. Ehinger, A. Oliva, and A. Torralba, "Recognizing scene viewpoint using panoramic place representation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012.

[8] Q. Tian and J. J. Clark, "Real-time specularity detection using unnormalized wiener entropy," in *Computer and Robot Vision (CRV), 2013 International Conference on*. IEEE, 2013, pp. 356–363.

[9] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," ser. SIGGRAPH '01. ACM, 2001.

[10] P.-P. Sloan, J. Kautz, and J. Snyder, "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments," ser. SIGGRAPH 2002. ACM, 2002.

[11] R. Ng, R. Ramamoorthi, and P. Hanrahan, "All-frequency shadows using non-linear wavelet lighting approximation," in *Proc. of ACM SIGGRAPH 2003*. ACM, 2003.

[12] D. Nowrouzezahrai, S. Geiger, K. Mitchell, R. Sumner, W. Jarosz, and M. Gross, "Light factorization for mixed-frequency shadows in augmented reality," in *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*. Washington, DC, USA: IEEE Computer Society, 2011.

[13] B. Tunwattanapong, A. Ghosh, and P. Debevec, "Practical image-based relighting and editing with spherical-harmonics and local lights," in *Visual Media Production (CVMP), 2011 Conference for*. IEEE, 2011, pp. 138–147.

[14] P. Debevec, "A median cut algorithm for light probe sampling," in *ACM SIGGRAPH 2006 Courses*. ACM, 2006.

[15] K. Viriyothai and P. Debevec, "Variance minimization light probe sampling," in *SIGGRAPH'09: Posters*. ACM, 2009, p. 92.

[16] S. Agarwal, R. Ramamoorthi, S. Belongie, and H. W. Jensen, "Structured importance sampling of environment maps," *ACM Transaction on Graphics*, 2003.

[17] H. Lu, R. Pacanowski, and X. Granier, "Real-time importance sampling of dynamic environment maps," in *EG 2013 - Short Papers*, 2013.

[18] M. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J. Lalonde, "Learning to predict indoor illumination from a single image," *ACM Transactions on Graphics (SIGGRAPH Asia)*, 2017.

[19] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination from shadows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 290–300, March 2003.

[20] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proceedings of SIGGRAPH '98*. ACM, 1998, pp. 189–198.

[21] M. K. Stern and J. H. Johnson, "Just noticeable difference," *Corsini Encyclopedia of Psychology*, 2010.

[22] G. Mather, *Essentials of sensation and perception*. Routledge, 2014.

[23] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[24] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural similarity metrics for texture analysis and retrieval," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 2225–2228.

[25] X. Zhao, M. G. Reyes, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for retrieval applications," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 1196–1199.

[26] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Perceptual similarity metrics for retrieval of natural textures," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*. IEEE, 2009, pp. 1–5.

[27] A. Chalmers, J. Zhao, D. Medeiros, and T. Rhee, "Reconstructing reflection maps using a stacked-cnn for mixed reality rendering," *IEEE Transactions on Visualization and Computer Graphics*, 2020.