Automatic Estimation of Illumination Features for Indoor Photorealistic Rendering in Augmented Reality

Hasan Alhajhamad$^{(\boxtimes)}$, Mohd Shahrizal Sunar, and Hoshang Kolivand

Media and Game Innovation Centre of Excellence (MaGIC-X) UTM-IRDA,
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
{hzhamad,kolivand}@magicx.my, shahrizal@utm.my
http://magicx.utm.my

Abstract. In this paper, a fast and practical algorithm is presented to estimate the multiple number of lights from every single indoor scene image in Augmented Reality environmet. This algorithm provides a way to accurately estimate the position, directions, and intensities properties of the light sources in a scene. Unlike other state-of-the-art algorithms, it is able to give accurate results without any essential analysis on the objects in the scene. It uses the analysis of the saturation channel HSV data. The evaluation is done by testing a ground truth dataset of synthetic and real images with known properties of lights and then comparing the results with other studies in the field.

Keywords: Illumination estimation · Shadow detection · Photorealistic augmented reality · Indoor spatial images · Computer vision

1 Introduction

Undoubtedly, the estimation of light illumination is one of the trickiest tasks in computer vision especially for indoor scenes. Presence of multiple light sources of different sizes and shapes, intensities and spectral features is a typical condition for such environments. The image based lighting is relatively the latest advanced approach for modeling light which achieves high quality results yet at the cost of processing time [1]. However, the requirements of complex hardware setup with additional cameras and/or light probes based on highly dynamic and superior image resolution are the main shortcoming of this approach.

Currently, the demands in the context of augmented reality applications are exponentially growing, where researchers and engineers have dedicated enormous efforts. Lately, the feasibility to augment real scenes with arbitrary objects and animations opened up broad prospects in the areas of design, entertainment, and human-computer interaction. Therefore, correct estimation of lighting conditions such as 3D positions and colors inside the scene appears to be a crucial step in creating the rendering highly realistic and convincing.

A modified approach that allows direct estimation of the positions of light sources is proposed but it also uses cumbersome hardware [2]. The most popular alternatives to image-based lighting approaches usually aim on the detection and direct analysis of shadings. Generally, these techniques are more suitable for outdoor environments with strong casts, directed light sources, and simple geometry. An exhaustive survey of the cast detection methods is provided in [3], while the possibility of their integration in real-time augmented reality systems is reviewed in [4,5].

Emergent technologies and vision-based robotics for fine-tuning the digital images suffer from lighting factors in real images [6,7]. Information regarding the light source distribution in an image facilitates in analyzing the objects, shadows, and noises present in the scene. Initial dismissal of detailed information related to light sources results inherent difficulty in analyzing the scene in the indoor environments [8].

Appearances of various lighting sources are found to be the major difficulties in real-time image processing. Detection of light sources permits one to estimate their position, direction, and intensity [9–11]. Figure 1 illustrates the background setting of indoor and outdoor scene illumination.
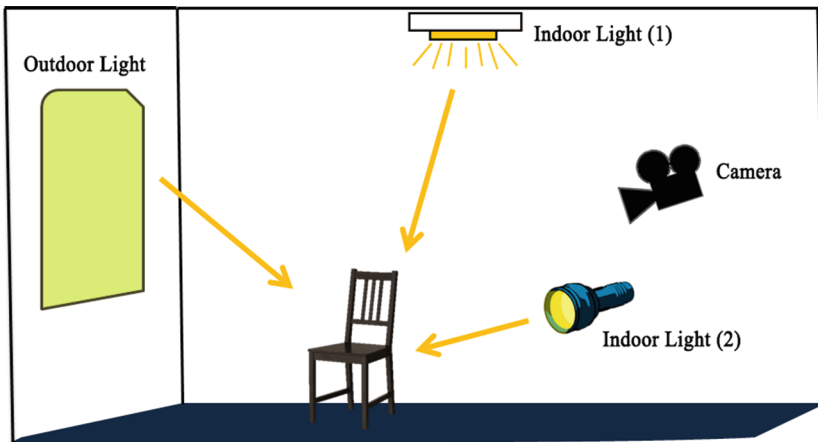

Fig. 1. Setting of indoor and outdoor scene illumination.

In this work, we generate the information on the localization of light sources in real scenes. Light source detection involves different techniques including probe detection which extracts the light properties from a known 3D object. Nevertheless, these techniques are not applicable to multiple lights and require the proper understanding of all 3D objects present in the real scene. Thus, we obtain the information on lighting without accessing any object in the scene itself. Indoor digital images possess varying numbers and directions of light. Researchers aim to estimate the approximate number of light sources to recognize the actual cause of occlusion which may appear from an object, shadow

or noise [7,12–15]. Therefore, the nine directions constriction of every light source in a 2D image is considered.

The presence of high diffuse sunlight in outdoor scenes makes their geometry much wider with less disparity lighting levels. Alternatively, indoor scenes suffer from limited geometry and higher disparity lighting levels due to the presence of multiple lights with different properties and influences. In fact, clear differentiation between these lights and their properties in the scene are extremely difficult. Consequently, we focus only on the color, intensity, position, and direction of each light source.

The algorithm is evaluated based on its accuracy and performance. The accuracy is tested using Lopez-Moreno evaluation where the results are compared with Light Source Detection (LSD) algorithm and synthetic image of known lights. Meanwhile, the performance of LSD algorithm is based on the number of frames per second. It is hoped that our systematic assessment on the automatic estimation of illumination characteristics may improve the performance of sundry applications including real-time digital photography editing, interactive photorealistic augmented reality, vision-based robotics, medical images, or New Media Art.

2   Related Work

Previous researches on the light detection are mostly based on 2D images in Computer Vision. They concerned with the object based and statistical based techniques. Recently, many methods are introduced [13,15–17] since the founding work of Pentland [6]. Majority of them either focused on extracting the properties of light by receiving information on the image objects or on extorting one direction of light. Fewer methods exploring the shadow detection obtain information on the light sources [18,19].

Strong assumptions on the background environment are needed for light source detection. Some of them analyze the objects and others estimate the geometry. Recovery of the light via object contours analysis is developed. Light sources illumination are estimated by inserting a hemisphere to the environment and analyzing the reflectance of light on it [20]. Many of them are based on the analyses of the reflectance of lights on real objects' contours in the image [16]. Shadow detection techniques are proven to be satisfactory in tackling the illumination issues [21–24]. Furthermore, recent works on the object analysis methods achieved relatively accurate results. Despite intensive efforts the precise detection of objects in the scene requiring the information of light sources is far from being achieved.

In the Augmented Reality, the first illumination estimation model was based on a simple point light source without ambient light assuming uniformly colored objects. The distant point light source is defined by an illumination direction [13]. Then, Kanbara [25] used a marker with a mirror ball to resolve geometric and photometric registration. Two methods rely on the illuminated geometry with no requirement for knowledge of a specific calibration object, one method

was proposed by Sato from shadows [26] and another from Shading by Wang [17]. Then, Wang [20] integrated the above two methods and proposed a new method for multiple directional source estimation. The model came after had used omni-directional stereo cameras to capture environment illumination and to model the distant parts of the environment [26]. Stumpfel [27] introduced other approach by photographed mirror spheres to capture the illuminants in the scene. This paper discusses another way to realize the illuminants in real scene in real-time.

From the get-go, most of the mentioned methods either force constraints on the scene, require more information for detection like depth sensors, or non real-time like in image relighting,

3   Light Source Detection Method

This paper attempts to extract multiple light sources from a single indoor scene image in the absence of any constraint on object detection. The information is directly acquired from the scene to develop a constraint free detection method. It is customary to introduce the structure of the light source estimation step-by-step as shown in Fig. 2. Firstly, a camera is placed in the indoor environment. Then, the light image is extracted from a single RGB image to extract the lighted areas of the scene. Synchronously, a shadow image is extracted from the same image to detect the shadowed areas of the scene. Finally, the calculation of the illumination properties is performed based on the relation between the lighted and shadowed areas as described in the following sections.

3.1   Illustration of Light Image

Even though indoor RGB images are captured by high quality cameras, however the presence of light, noise and scene variations remain a problem to be solved.
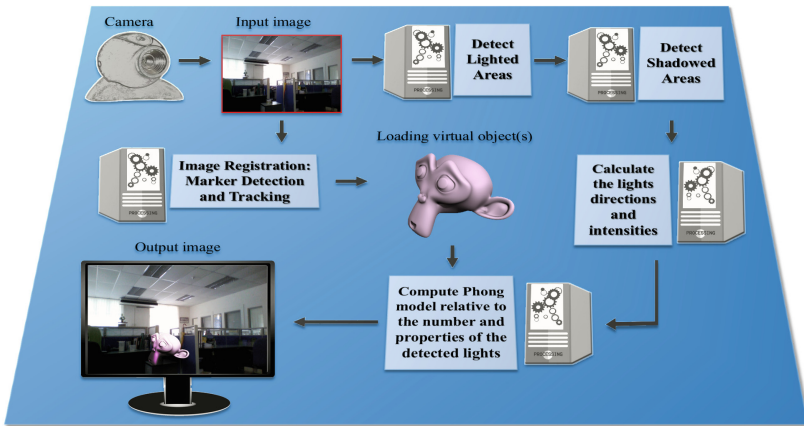

Fig. 2. The pipeline of the illumination sources detection model.

Consequently, the user must extract less noise and get more illumination stability while capturing. The first objective is therefore to control the illumination in the image. Supposing that the output is a single channel light image, one in interested to extract lightness levels of each pixel in the RGB image. Let I_{RGB} be the input of 8-bits RGB image, n is the total number of pixels in the image and ξ_{lights} is the set of the resulted lights. Here each light represents the area of the direct or indirect real light in I_{RGB}. Then, I_{RGB} and ξ_{lights} can be defined as,

$$I_{RGB} = \begin{pmatrix} R_1, G_1, B_1 \\ . \\ . \\ . \\ R_n, G_n, B_n \end{pmatrix}$$

$$\xi_{lights} = \begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ ... \\ C_{nl} \end{pmatrix}$$

where, R_i, G_i and B_i are the ith red, green and blue channels, respectively. C_i consists of the area $C_i(\text{area})$ of a detected light in some position $C_i(\text{pos})$, direction $C_i(\text{dir})$, and intensity C_i (int), and nl is the number of detected lights. C_i is expressed as,

$$C_i = \begin{pmatrix} area \\ pos \\ dir \\ int \end{pmatrix}$$

To compute each element C_i, the following algorithm comprised of six major steps is proposed:

Algorithm 1. Light Source Detection

1: Convert I_{RGB} to HSL format image I_{hsl}
2: Split I_{hsl} to its Hue, Saturation, and Lightness
3: Calculate Otsu's threshold on the Saturation channel using look-up table
4: Filter the I_{hsl} using Gaussian blur
5: Extract the set of contours ξ_{lights} for the detected spots in I_{hsl}
6: Represent the position $C(\text{pos})$ of each element of ξ_{lights} as the center of the contour C

The red-green-blue (RGB) to Hue-saturation-lightness (HSL) conversion formula is common. The RGB values of one pixel range from 0 to 255 for each channel. Similarly, the value of each channel in one HSL pixel lies in the range

of 0 to 1. Consequently, the conversion value of the result is measured by dividing the I_{RGB} channel value by 255 as hereunder:

Let us assume,

$$I_{HSL} = \begin{pmatrix} Hue_1, Sat_1, Lig_1 \\ . \\ . \\ . \\ Hue_n, Sat_n, Lig_n \end{pmatrix}$$

then,

$$R' \leftarrow R/255$$
$$G' \leftarrow G/255$$
$$B' \leftarrow B/255$$

Assuming that C_{min} and C_{max} are the minimum and maximum value of each channel in one pixel, then the difference $diff$ is calculated from,

$$C_{max} = \max(R', G', B')$$
$$C_{min} = \min(R', G', B')$$
$$diff = C_{max} \text{ - } C_{min}$$

Accordingly,

$$H = \begin{cases} 60° \times (\frac{G'-B'}{diff} \bmod 6), & C_{max} = R' \\ 60° \times (\frac{B'-R'}{diff} + 2), & C_{max} = G' \\ 60° \times (\frac{R'-G'}{diff} + 4), & C_{max} = B' \end{cases}$$

$$L = \frac{C_{max} + C_{min}}{2}$$

$$S = \begin{cases} 0, & 0 \\ \frac{diff}{1-|(2 \times Lightness)-1|}, & otherwise \end{cases}$$

Look-up table is calculated and the optimal performance is obtained to achieve the values of each pixel. Applying Otsu's thresholding [28] on the Saturation channel image a binary image is obtained. The lighted areas of the scene is represented by white spots in the binary image and remained black elsewhere (Fig. 3). By blurring the extracted binary image, the contours of the lighted areas enhanced the smoothness approximation of light which provided better detection quality.
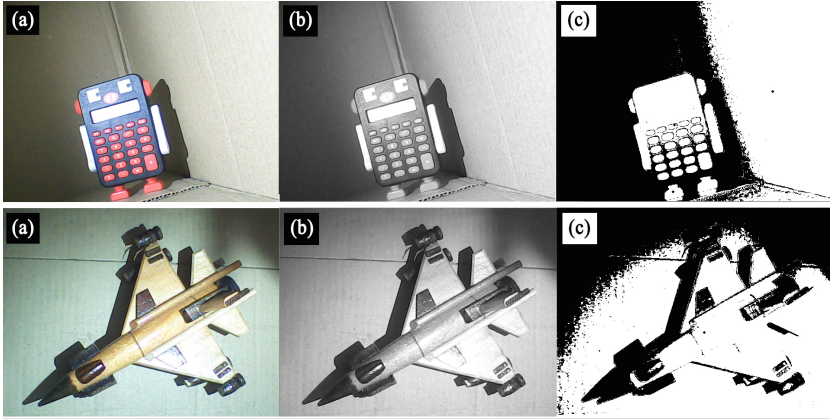
Fig. 3. (a) Original images (b) single channel saturation images (c) lighted areas in binary images.eps

3.2 Shadows Detection

After identifying the albedo spaces, contour detection is implemented to extract the covered areas of the light sources from the binary image. In this case filled closed contours representing each light source area are drawn. Contours are measured by following Suzuki and Abe algorithm [29]. Subsequently, the resulted contour areas are stored in moments. Each moment M_i consists of contour area and its mass center (C_i). In case a moment area covers a small scale in the image (e.g. 30 pixels), the moment is discarded. If the moment is in large-scale area (e.g. $\frac{n}{2}$ or more) then it is segmented into fixed or automatic number of moments. The division processes fixed by 2 moments yields,

$$M_i = \begin{cases} discard, & M_{area} < 30 \\ <M_{area}, C_i>, & \frac{n}{2} > M_{area} \geq 30 \\ Segment(M_i, 2), & M_{area} \geq \frac{n}{2} \end{cases}$$

Thus, the achieved result is a set of moments represented by their mass centers C_i each to behave like a source or a reflection of light as displayed in Fig. 3.

Algorithm 2. Shadows Detection

1: Calculate Otsu's threshold of the dark colors from the Saturation channel using look-up table
2: Filter the I_{hsl} using Gaussian blur
3: Extract the set of contours $\xi_{shadows}$ for the detected spots in I_{hsl}
4: Represent each element of $\xi_{shadows}$ as the center of the contour D

Definitely, the quality of results entirely depends on the proper detection of each illuminant. Illuminant direction is detected by estimating the differences between light sources and shadows in the scene. Estimating shadows in the scene does not essentially mean that only shadowed spaces are detected it also include dark objects. This is because the presence of dark objects in the scene behave the same way as shadows vis-à-vis its reflectance properties.

Identification of the shadow behavior in the saturation image clearly reveals the relation between the light source position, occlusion, and shadow. The results show that if the mass center of the detected light source is obtained and the mass center of the shadow is taken in the same way, the direction of the light forms the marching vector from the light towards the mass center of the shadow moment. Algorithm 2 explains the steps that are adopted in Light Source Detection in Algorithm 1.
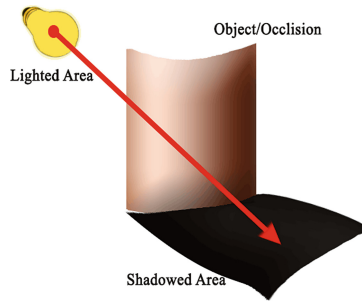

Fig. 4. Direction of the light source starting from its mass center to that of shadow centre for the occluded object.

Figure 4 shows the method to resolve the direction of one light source to one detected shadow. The problem remains in the detection of large number of lights and shadows in the scene. The distribution of lights and the relation between them and other shadows in the scene can lead to approximate both the direction and intensity of each light.

3.3   Light Source Direction and Intensity Detection

This problem is overcome by taking each light and calculating the average distance between its mass center and all other mass centers shadows in the scene. The near and far angles for each shadow are considered. In case some angles are close to each other, the intensity of the light towards this direction must be less than the farther ones. For some far angles the intensity of the light is higher. Algorithm 3 shows the steps for computing the multiple light intensities and light directions.

The algorithm simply considers each of the detected light sources and analyzes the distances between them and other shadows in the scene. Only the close lights and shadows are included.

Algorithm 3. Calculation of multiple lights directions and intensities

---

1: Define $C_i(int)$ as 0.5 for all C_i in ξ_{lights}.

2: For each light mass center C_i in ξ_{lights} , execute step 2.

3: For each shadow mass center D_j in $\xi_{shadows}$, execute step 3 and 4.

4: calculate the average distance between C_i and D_j.

5: calculate the C_i angle between each shadow. If some shadows are have nearly equal angles, reduce the intensity value of C_i otherwise increase the value.

---

4   Implementation

A data set of several images is gathered to test the accuracy and performance of the proposed algorithm. Multi-camera Multi-lighting Dome (MCML) Data Sets are used in [30]. These datasets are image sequences captured in a studio for 20 views evenly paced on a ring inside a lighting dome. All images in the dataset are captured from a single PC camera that supports HD quality images.

The experiments are performed in three different scenarios. The first scenario uses a single light source in a known position, direction, and intensity placed in the dark area. Figure 5 depicts the steps to estimate the illumination of a single light source. The lighted areas image in Fig. 5 shows the light and the reflection of it as dark spots which reveals the mass center of them as a white marker based on Algorithm 1.
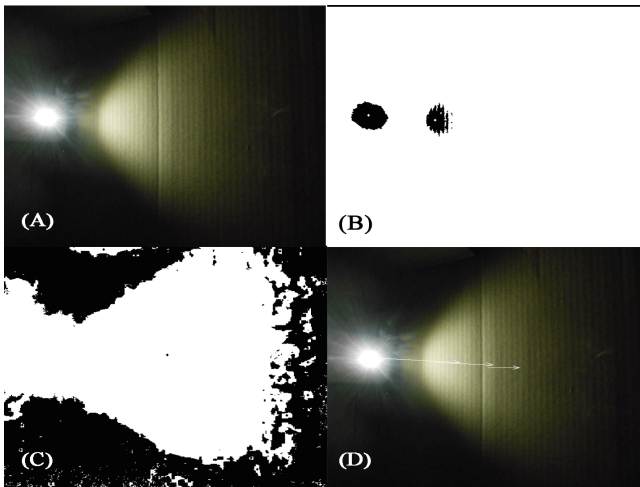

Fig. 5. Scenarios of (a) original image (b) lighted areas image (c) shadowed areas image (d) resulted image.

The shadowed areas image displays the shadow spaces in the image based on Algorithm 2 by considering the measurement of the mass centers of each

detected shadow area. Each estimated light is exemplified by a white arrow to represent the accuracy of the data. The single arrow's starting point represents the position of the light source where its length signifies the intensity of the light and the head of the arrow indicates its direction.
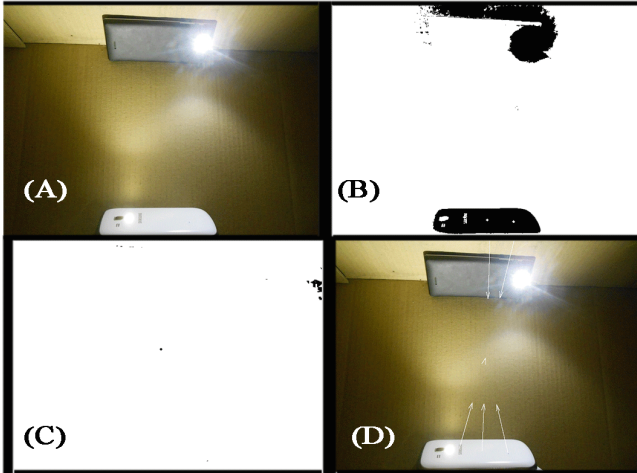

Fig. 6. (a) Original image (b) lighted areas image (c) shadowed areas image (d) resulted image.

In the second scenario, two light sources were placed in a dark space. Figure 6 illustrates the same steps as performed in the first experiment. The results successfully prove the correct position, direction, and intensities of both light sources. The original image exhibits that the two light sources are situated to face each other in a dark area and the lighted areas image estimated total six elements from the original image. Each light is represented as a dark spot, where only two shadowed elements are detected. However, the arrows in the resulted image outnumbered the actual number of light sources due to the existence of reflections.

Figure 7 illustrates the snapshot of an arbitrary indoor scene where the size of the room is considered to be quite large. The estimated number of light sources is nine and the number of detected shadowed areas is fifty four. The resulted image shows that the algorithm had successfully obtained the real light sources in the scene from the image.

Accurately, speed improvement is a major contribution of this paper. For 1024×768 image, the mean processing time is found to vary between 40–60 ms on a laptop with Intel® Core™ i7-4702MQ CPU at 2.20 GHz (8 cores) and 4 GB RAM. The algorithm implementation is administered under in-hand system specifications.
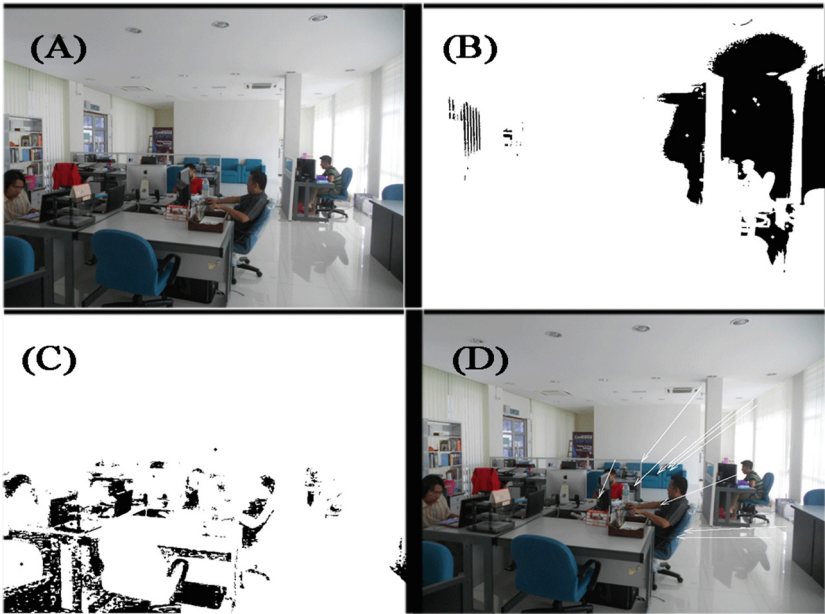
Fig. 7. (a) Original image (b) lighted areas image (c) shadowed areas image (d) resulted image.

Table 1 summarizes the speed of the algorithm in each of the previously described experiments. The estimated performances for each experiment do not exhibit as such big difference between single, two, or multiple light sources.

Table 1. Performance of the algorithm.

Experiment	Frame rate	
	Minimum	Maximum
Single light source	45	54
Two light sources	42	50
Complete indoor scene	40	49

Our results show an improvement from the Multi View Stereo (MVS) algorithms perspective when compared with the work of Wu et al. [30]. This algorithm is advantageous because it overcomes the case of non-Lambertian objects compare to Wu et al. Furthermore, their method requires the access of the scene and the construction of a mesh which leads to slow performance.

Figure 8 depicts the use of similar dataset on the virtual model data with the difference in results.
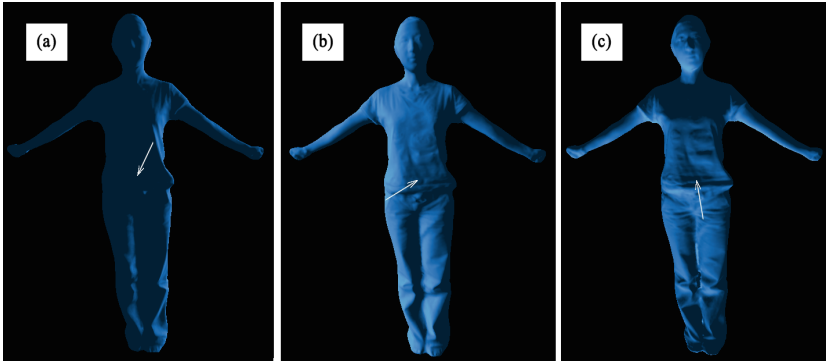
Fig. 8. Implementation on the Wu et al. datasets (a) position, (b) direction, and (c) intensity of single light source in different poses projected on a synthetic object.

The position, direction, and intensity of a single light source in different poses projected on a synthetic object are implemented. It produces impressive results with high level accuracy. The arrows position in the middle of the scene confronting the object. Furthermore, the arrow is able to follow the real light direction and for its intensity immediately after changing the direction. In comparison to our method the earlier one [30] is unable to detect the real light source, and only useful for implementing the relighting on the scene. We assert that the proposed method is efficiently capable in achieving information instantly with the frequent change of the light position, direction, or intensity.
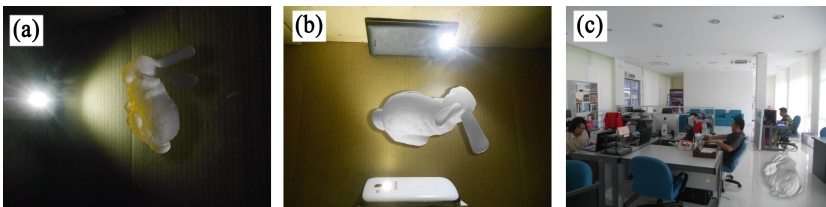

Fig. 9. Implementation of the results of the algorithm in AR scene (a) positioning a single light source, (b) two lights, and (c) complex scene with multiple light source and their projection on a synthetic object.eps

As a result of our previous experiments, Fig. 9 shows an AR scene where a composition of the real scene and a virtual object in real-time after performing the steps in Fig. 3. The provided images where captured in three illumination states; with a single light source, two lights, and complex scene with multiple lights.

5   Conclusions

From the previous work, a novel algorithm for detecting illumination in a scene was described. The algorithm estimates the lights and the reflected lights without going into detailed analysis. Experiments were done by comparing real and synthetic datasets. Real time experiments were done using a conventional camera. The comparability of the performance was found less complex than other image processing techniques. The algorithm has two advantages: it measures the illumination in the scene without access to the analysis of the scene and it provides information about the positions, directions and intensities of the detected illuminants. The confidence of the algorithm's results can vary dependently on the size of the room.

References

1. Debevec, P.: Image-based lighting. IEEE Comput. Graph. Appl. **22**(2), 26–34 (2002)
2. Frahm, J.-M., Koeser, K., Grest, D.: Markerless augmented reality with light source estimation for direct illumination. In: Conference on Visual Media Production CVMP, London, pp. 211–220. IET (2005)
3. Al-Najdawi, N., Bez, H.: A survey of cast shadow detection algorithms. Pattern Recogn. Lett. **33**(6), 752–764 (2012)
4. Jacobs, K., Loscos, C.: Classification of illumination methods for mixed reality. Comput. Graph. Forum **25**(1), 29–51 (2004)
5. Neverova, N., Muselet, D., Trémeau, A.: Lighting estimation in indoor environments from low-quality images. In: ECCV 2012 Proceedings of the 12th International Conference on Computer Vision, pp. 380–389 (2012)
6. Pentland, A.P.: Finding the illuminant direction. J. Opt. Soc. Am. **72**(4), 448–455 (1982)
7. Yeoh, R.C., Zhou, S.Z.: Consistent real-time lighting for virtual objects in augmented reality. In: 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 223–224. IEEE (2009)
8. Bingham, M.: An Interest Point Based Illumination Condition Matching Approach to Photometric Registration Within Augmented Reality Worlds (2011)
9. Bouganis, C.S., Brookes, M.: Statistical multiple light source detection. IET Comput. Vis. **1**(2), 79–91 (2007)
10. Lopez-moreno, J., Hadap, S., Reinhard, E., Gutierrez, D.: Light source detection in photographs. In: Andujar, C., Lluch, J. (eds.) Congreso Espanol de Informatica Grafica, vol. 11, pp. 161–168. Eurographics S.E. (2009)
11. Wei, J.: Robust recovery of multiple light source based on local light source constant constraint. Pattern Recogn. Lett. **24**(1–3), 159–172 (2003)
12. Agusanto, K., Li, L., Chuangui, Z., Sing, N.W.: Photorealistic rendering for augmented reality using environment illumination. In: The Second IEEE and ACM International Symposium on Mixed and Augmented Reality 2003 Proceedings, vol. 3, pp. 208–216. IEEE Computer Society (2003)

13. Zheng, Q., Chellappa, R.: Estimation of illuminant direction, albedo, and shape from shading. IEEE Trans. Pattern Anal. Mach. Intell. **13**(7), 680–702 (1991)
14. Zhang, Y., Yang, Y.-H.: Multiple illuminant direction detection with application to image synthesis. IEEE Trans. Pattern Anal. Mach. Intell. **23**(8), 915–920 (2001)
15. Lopez-Moreno, J., Hadap, S., Reinhard, E., Gutierrez, D.: Compositing images through light source detection. Comput. Graph. **34**(6), 698–707 (2010)
16. Lopez-Moreno, J., Garces, E., Hadap, S., Reinhard, E., Gutierrez, D.: Multiple light source estimation in a single image. Comput. Graph. Forum **32**(8), 170–182 (2013)
17. Wang, Y., Samaras, D.: Estimation of multiple illuminants from a single image of arbitrary known geometry. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 272–288. Springer, Heidelberg (2002)
18. Noh, Z., Sunar, M.S.: Soft shadow rendering based on real light source estimation in augmented reality. Adv. Multimedia - Int. J. **1**(2), 26–36 (2010)
19. Noh, Z., Sunar, M.S.: A review of shadow techniques in augmented reality. In: Second International Conference on Machine Vision, pp. 320–324. IEEE (2009)
20. Wang, Y., Samaras, D.: Estimation of multiple directional light sources for synthesis of augmented reality images. Graph. Models **65**(4), 185–205 (2003)
21. Panagopoulos, A., Wang, C., Samaras, D., Paragios, N.: Illumination estimation and cast shadow detection through a higher-order graphical model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 673–680. Image Analysis Lab, Computer Science Dept., Stony Brook University, IEEE, NY, USA (2011)
22. Mei, X., Ling, H., Jacobs, D.W.: Illumination recovery from image with cast shadows via sparse representation. IEEE Trans. Image Process. : Publ. IEEE Signal Process. Soci. **20**(8), 2366–2377 (2011)
23. Sato, I., Sato, Y., Ikeuchi, K.: Illumination from shadows. IEEE Trans. Pattern Anal. Mach. Intell. **25**(3), 290–300 (2003)
24. Panagopoulos, A., Wang, C., Samaras, D., Paragios, N.: Estimating shadows with the bright channel cue. In: Kutulakos, K.N. (ed.) ECCV 2010 Workshops, Part II. LNCS, vol. 6554, pp. 1–12. Springer, Heidelberg (2012)
25. Kanbara, M., Yokoya, N.: Real-time estimation of light source environment for photorealistic augmented reality. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2, pp. 911–914. IEEE (2004)
26. Sato, I., Sato, Y., Ikeuchi, K.: Acquiring a radiance distribution to superimpose virtual objects onto a real scene. IEEE Trans. Visual Comput. Graphics **5**(1), 1–12 (1999)
27. Stumpfel, J., Jones, A., Wenger, A., Tchou, C., Hawkins, T., Debevec, P.: Direct HDR capture of the sun and sky. In: ACM SIGGRAPH 2006 Courses on - SIGGRAPH 2006. AFRIGRAPH 2004, vol. 1, p. 5. ACM Press (2006)
28. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst., Man, Cybern. **9**(1), 62–66 (1979)
29. Suzuki, S., Be, K.: Topological structural analysis of digitized binary images by border following. Comput. Vis., Graph., Image Process. **30**(1), 32–46 (1985)
30. Wu, C., Liu, Y., Dai, Q., Wilburn, B.: Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. IEEE Trans. Vis. Comput. Graph. **17**(8), 1082–1095 (2011)