Real-time Illumination and Visual Coherence for Photorealistic Augmented/Mixed Reality

A'AESHAH ALHAKAMY, Indiana University-Purdue University, Indianapolis, USA and University of Tabuk, Saudi Arabia
MIHRAN TUCERYAN, Indiana University-Purdue University, Indianapolis, USA

A realistically inserted virtual object in the real-time physical environment is a desirable feature in augmented reality (AR) applications and mixed reality (MR) in general. This problem is considered a vital research area in computer graphics, a field that is experiencing ongoing discovery. The algorithms and methods used to obtain dynamic and real-time illumination measurement, estimating, and rendering of augmented reality scenes are utilized in many applications to achieve a realistic perception by humans. We cannot deny the powerful impact of the continuous development of computer vision and machine learning techniques accompanied by the original computer graphics and image processing methods to provide a significant range of novel AR/MR techniques. These techniques include methods for light source acquisition through image-based lighting or sampling, registering and estimating the lighting conditions, and composition of global illumination. In this review, we discussed the pipeline stages with the details elaborated about the methods and techniques that contributed to the development of providing a photo-realistic rendering, visual coherence, and interactive real-time illumination results in AR/MR.

49

CCS Concepts: • **Computing methodologies → Computer graphics**; **Graphics systems and interfaces**; Mixed/augmented reality; Perception;

Additional Key Words and Phrases: Augmented reality, mixed reality, visual coherence, real time, illumination, lighting condition, image-based lighting, reflectance, and shading, photo-realistic

ACM Reference format:
A'aeshah Alhakamy and Mihran Tuceryan. 2020. Real-time Illumination and Visual Coherence for Photorealistic Augmented/Mixed Reality. *ACM Comput. Surv.* 53, 3, Article 49 (May 2020), 34 pages.
https://doi.org/10.1145/3386496

1 INTRODUCTION

As is known, virtual reality (VR) provides a totally computer-generated environment for interaction with the user, while augmented reality (AR) embeds virtual contents that are registered directly to the physical environment. Augmented reality links the gap between the virtual and real worlds in both spatial and cognitive substance. In AR applications, the user's perception of

Authors' addresses: A. Alhakamy and M. Tuceryan, Indiana University-Purdue University Indianapolis, Department of Computer & Information Science, 723 W. Michigan St, SL280, Indianapolis, Indiana, USA; emails: aalhakam@iu.edu, aalhakami@ut.edu.sa, tuceryan@iu.edu.

0360-0300/2020/05-ART49 $15.00
https://doi.org/10.1145/3386496

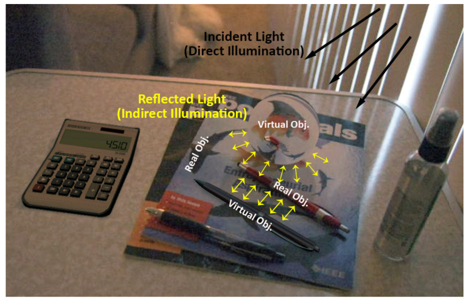

Fig. 1. Illustration of the light interaction between each type of objects; virtual and real [6].

the digital information is integrated and perceived as part of the real world. Furthermore, the Mixed Reality (MR) also refers to a hybrid reality for producing an innovative visualization and environments where the digital and physical objects interact and co-exist in real time.

Some of the previous works were defined without output devices specification such as head-mounted displays (HMD) [77], mobile devices, and cameras to limit the AR visual media. Thus, it may be difficult to realize that other other information were captured from the surrounding environment including audio, haptic, olfactory, or gustatory AR. It entails spatial registration and real-time control, which means precise alignment in three dimensions (3D) in real time for corresponding real and virtual information. The user of the AR system under this mandate could use at least some viewpoint control interactively, and the display of computer-registered augmentations will remain registered in the environment for each referenced object.

Many AR/MR applications nowadays can produce fair results, but they still lack the realistic output in which the human eye could be easily deceived and not recognize the virtual objects from the real ones. The ability to synthesize realistic images and integrate virtual objects flawlessly into physical environment scenes is one of the significant objectives of many AR systems. The ability to produce a photo-realistic augmented object depicting the real world illumination characteristic such as incident light, reflection, shading, and cast shadows, is becoming increasingly important.

Therefore, a large number of techniques were developed for capturing and estimating the light source and its conditions in the physical world. Also, many techniques were developed to augment objects in the right position. Computer vision, machine learning, and high dynamic range imaging (HDR) techniques expedited a variety of novel estimating and rendering methods. This article aims to provide a comprehensive review and comparison of capturing and estimating illumination techniques for augmented reality. For further reading, the reader is referred to the book by Schmalstieg and Hoellerer [94] where most of the concepts and methods mentioned in this article are described in detail.

Accurate registration and calibration for a consistent geometry was the focal point in the early research work, but which omits the effects of local or global illumination and cast shadows among virtual and real objects. Ignoring the impact of illumination effects during the augmentation of the scene resulted in producing poor-quality and sometimes visually confusing outputs. The current studies have shown that one of the critical elements for virtual objects to be perceived as realistic objects in real scenes is their photo-realistic rendering with an illumination consistent with the physical scene [11, 35, 47, 88, 94].

The idea is not only to illuminate the virtual objects with captured or estimated incident light in the real environment, but to further simulate cast shadows and reflected light interactions as a unified illumination among the different types of objects in the scene, whether they are virtual or real, see Figure 1. In this review, we provide a survey about multiple methods that accomplished illumination effects of a dynamic scene in different stages for more visual coherence in real time.

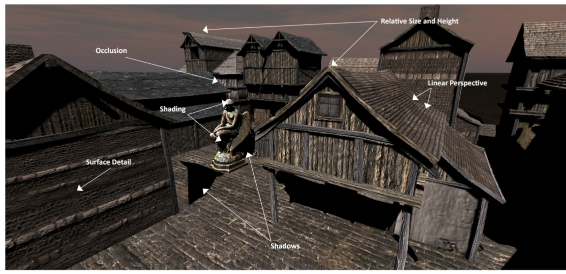

Fig. 2. Monocular depth cues allow scene structure interpretation from a single image.

We review methods on the sequence of familiar stages of mixed reality for acquisition, registration (estimation), composition and display to obtain a consistent illumination that includes the following points:

- **Capturing** the real-world light source using typical measurements using, for instance, omnidirectional HDR images.
- **Estimating** the light environment from images and video directly.
- **Compositing** the dynamic radiometric information from virtual to a real object such as shadow casting.
- **Rendering** efficient techniques in augmented reality using Monte Carlo ray tracing, precomputed radiance transport, and differential rendering.

For a more organized and focused survey, we are omitting the geometric calibration and registration aspects of this topic. Therefore, we only cover methods that involve illumination and lighting. However, the reader is referred to Reference [94] for more details. We include recent advances in studies and research from computer vision, computer graphics, and virtual/augmented reality. Before we dive into those specific details, we need to explore a small overview of the visual coherence, rendering equation, and light transfer model.

2 VISUAL COHERENCE: AN OVERVIEW

The ability to embed a three-dimensional virtual object into an image of the real scene means that objects should be rendered from a virtual camera with internal and external parameters that correspond to the physical camera viewing the physical scene. Fundamental depth cues can be obtained with such a calibrated camera. Information about the depth provides interpretation of three-dimensional structure from the viewpoint of the camera. Depth cues can be categorized as monocular or binocular where there are around 15–20 different depth cues. Cues from a single image are known as monocular, see Figure 2, while cues depicted in a pair of images are called binocular [40, 54, 94].

At present, several AR displays use a monocular video see-through mode. In general, for AR, the most important cue is the depth that can be produced by computer graphics software. Also, there are other essential depth cues such as the following:

- **Relative size:** the distance between the objects and the observer. The further it is, the smaller the object appears.
- **Relative height:** how far the object is from the other objects and where their base is higher in the image.
- **Perspective:** the convergence of the parallel lines as the distance from the observer increases.

- **Surface detail:** objects closer to the observer have more texture gradient and more fine-grained surface detail.
- **Atmospheric attenuation:** while closer objects appear clearer, most-distant objects can be blurred due to atmospheric effects.
- **Occlusion:** in the screen space, closer objects obscure the further ones along the line of sight.
- **Shading:** according to the source of light, the objects are illuminated resulting in the shading of the surfaces due to geometry.
- **Shadow:** objects blocking the light cast shadow on other objects.

These cues are delivered by well-equipped three-dimensional computer graphics tools. Some of them are more straightforward to produce by a virtual camera registered geometrically with a real one, such as size, perspective, height, and surface details. Atmospheric attenuation concerns far-field outdoor AR. However, the other cues, especially occlusion, shading, and shadows, demand attention in the AR rendering process [40, 54, 94].

Combining the real and virtual worlds in AR/MR extends the conventional rendering process in the computer graphics pipeline to involve more steps. The video see-through pipeline is better suited in this article than an optical see-through pipeline, which consists of the following stages:

- **Acquisition:** obtain a model or a set of data from the real scene such as geometry, materials, and illumination.
- **Registration:** transform the obtained sets of data, which is the standard photometric and geometric properties of one coordinate system of the real and virtual scenes.
- **Compositing:** merge the virtual objects and the real physical environment into one single image scene.
- **Display:** provide the user with the composited image.

The rendering process in AR/MR obviously is more complicated than the standard pipeline in computer graphics. In AR, we have to deal with the virtual scene and the real scene simultaneously to provide geometric and photometric characteristics for both scenes [94]. In this report, we discuss each stage with elaborated details about the methods and technologies that contributed to the development of providing a photo-realistic rendering and dynamic real-time illumination results in AR/MR.

3 RENDERING EQUATIONS AND LIGHT MODELS

Writing code about a three-dimensional rendering engine acquaints one with many lighting models and general concepts and definitions such as albedo, ambient lights, specular reflections, and diffuse colors. The reflection model of the diffuse surface (Lambertian surface) is the most straightforward lighting model, also known as "dot product lighting" or "cosine rule." The intensity of each light source also known as RGB color reflected from a surface is given by the following equation:

$$I = sC \times \sum_{i=1}^{nlights} lI_i \times (N \cdot L_i), \tag{1}$$

where sC is the color of the surface and lI represents the color of the light among multiple number of lights. The first fragment of the code is to calculate the incoming light from different directions, scale it by the cosine of the angle between surface normal (N) and light source direction (L), and then multiply the outcome by the reflection function for the diffuse surface that is a constant color [30].

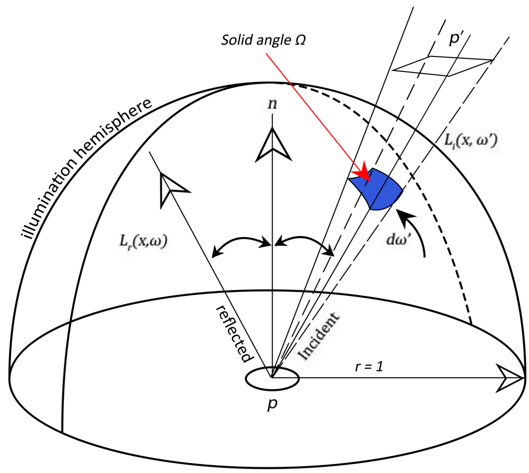

Fig. 3. Simple representation of the rendering equation.

This is the purest form of the rendering equation based on the physics to produce images in computer graphics. It is a standard concept where the entire realistic lighting must be measured. The rendering equation [45] in Figure 3 is represented as follows:

$$L_o(p, \overrightarrow{w}_o) = L_e(p, \overrightarrow{w}_o) + \int_S f_r(p, \overrightarrow{w}_i, \overrightarrow{w}_o) L(p', \overrightarrow{w}_i) G(p, p') V(p, p') dw_i, \tag{2}$$

where $L_o(p, \overrightarrow{w}_o)$ represents the output radiance or the intensity reflected from position p into the reflection direction $\overrightarrow{w}_o$. In the second part of the equation, $L_e(p, \overrightarrow{w}_o)$ refers to the emitted radiance/light from material at point x by the object itself. Then we have S, which describes the upper hemisphere surrounding the surface normal at the point p. The expression $f_r(p, \overrightarrow{w}_i, \overrightarrow{w}_o)$ represents the bidirectional reflectance distribution function of the surface (BRDF) at point p. Additionally, $L(p', \overrightarrow{w}_i)$ accounts for the incoming radiance at point p' arriving from all the directions $\overrightarrow{w}_i$ according to the BRDF and the surface normal. Also, $G(p, p')$ represents the geometric relationship between points p and p', and $V(p, p')$ is a visibility test for diminished intensity per unit area that returns 1 if p could see p' or returns 0 otherwise.

The only tricky concept while writing ray tracers is the usage of differential angles for representing a series of rays and the integral self-solving, where the background of global illumination is quite achievable [30]. Separating the incident radiance into different parts helps us comprehend the light transport in AR/MR where mixed scenes are involved:

$$L(p', \overrightarrow{w}_i) = L^r(p', \overrightarrow{w}_i) + L^v(p', \overrightarrow{w}_i) + L^{v,r}(p', \overrightarrow{w}_i). \tag{3}$$

The first part of the incidence radiance is $L^r(p', \overrightarrow{w}_i)$, which refers to the real scene incident radiance that is not yet reflected on the surface of the virtual objects. The second part, $L^v(p', \overrightarrow{w}_i)$, represents the incident radiance that could be emitted from either the virtual and real objects then reflected once or multiple times on the surfaces of the virtual objects. The third part, $L^{v,r}(p', \overrightarrow{w}_i)$ stands for the incident radiance that already interacts with both real and virtual objects [60].

These three parts must be accounted for for a precise calculation of the radiance that is outgoing at a particular location on the virtual object. The first part could be sampled directly from the real scene by capturing or estimating the light conditions. The other parts are computed recursively utilizing the global illumination algorithms [82].

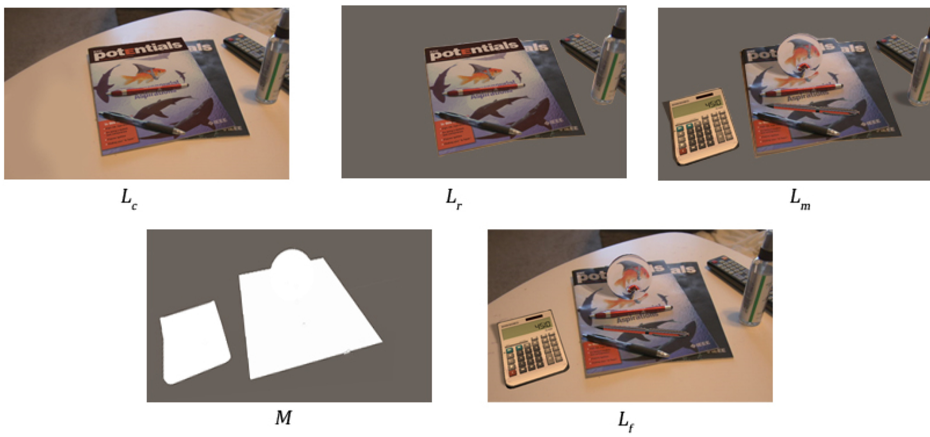

Fig. 4. Illustration of differential rendering [5, 6].

Additionally, the third part must mimic and influence the effects of the synthetic objects as if they were in the real scene. A model is required to describe the geometry of the real surfaces and reflectance properties impacted by the virtual objects. Therefore, usually dividing the real scene into two parts is the typical approach: a distant scene where the virtual objects are not affecting the scene, and a local scene where they are affecting the scene. The lighting simulation is enabled in the local scene model [18].

The interaction among the synthetic (virtual) objects and real scene or objects is computed using a standard technique called differential rendering. The method involves updating the pixels of the background image or video for applying the difference of the reflected radiance as:

$$L_f = M \odot L_m + (1 - M) \odot (L_c + L_m - L_r), \tag{4}$$

where L_f is the *final image* composited, M represents the rendering alpha *mask* of the first image created (1 if the pixels overlapped with a virtual object, or 0 if they overlapped with a real object). While L_m stands for the resulted *image* after rendering both real/virtual objects with *mixed radiance*, L_r is the *image* that represent the real objects rendered with the *real radiance*. Finally, L_c refers to the *source image* or video taken by the camera. The symbol $\odot$ represents the Hadamard product of the respective color vector [47, 60], see Figure 4.

4 ACQUISITION OF LIGHTING METHODS

This section explores multiple techniques that explicitly acquire the lighting conditions and illumination aspects from the real scene where the virtual objects are located using tracking algorithms.

4.1 Image-Based Lighting (IBL)

A single omnidirectional image known as an environment map was utilized to represent the incident radiance at a certain point of the real scene. This procedure captured the angular distribution at that single point. During rendering, the captured panoramic image, which is also called a light probe, is used to recreate the real physical lighting conditions on the virtual scene or objects. We can see in Figure 5 how to capture the scene source light while rendering using the panoramic HDR image.

Any pixel in an HDR environment map can be considered as a lighting incident measurement over an angle solidly subtended by the same pixel. The virtual object's coordinate system must be

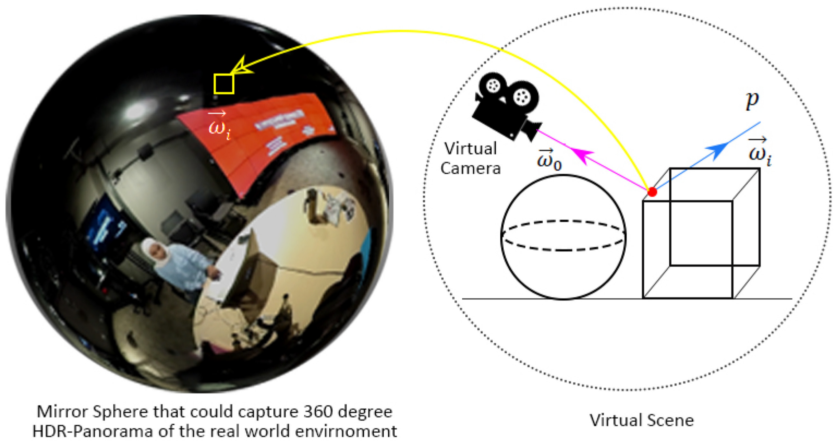

Fig. 5. Allocation in HDR environment map and spherical projection concept [3, 4, 60].

aligned with the environment map. This allows us to interpret a ray from a pixel not intersecting the virtual object that is aligned with coordinate system while sampling the lighting condition.

In earlier work, to simulate the perfect specular reflection and refraction, the panoramic images were used. At a surface point, the reflected or refracted vector could be used for specular scattering events as a direct lookup in the environment map. The method encounters some limitations, as it is able to handle only the specular effects, and also requires the environment map to be preprocessed with a low-pass kernel using the surface normal to lookup the reflected radiance.

The first person who proposed integrating arbitrary BRDFs was Debevec [18], who also incorporated global illumination effects and established the study of the relationship/interaction among the local parts of the real physical scene with the synthetic objects. He employed a HDR environment map for capturing the real scene incident light. The full range of light can be captured using the HDR imaging as linear response measurements [54]. The light conditions of the real world can be represented accurately by measuring or estimating the full dynamic range in the scene directly. For further reading of the HDR image capturing the reader is referred to References [24, 85].

The idea behind capturing omnidirectional HDR images was practical and straightforward. The method was based on placing a mirrored sphere where incident illumination needs to be obtained in the scene. Then, as many images as required are captured for the scene HDR photographs using a standard lens camera. The reason for taking multiple pictures is to cover all blind spots and to improve the resolution, because a single image includes only the area in front of the camera resulting in reduced resolution. However, this system has one problem, which is that the directions were measured from different points in space, i.e., there is no central projection point. For accomplishing this objective, some studies used parabolic or hyperbolic mirrors but were rarely used for real-life applications [2, 53]. Also, capturing omnidirectional HDR environment maps could be achieved using fish-eye lenses, specialized hardware and cameras, and panorama stitching [9, 11, 46, 47, 53, 86, 88, 95].

In an outdoor scene, the illumination conditions usually revealed an actual high dynamic range that makes it necessary to capture images of the sun and the sky using a number of techniques that utilize an accurate combination of aperture, exposure time, and filters [2, 35, 65, 108].

The ambient light is usually exceeded by many orders of magnitude in the HDR images, which are vital for capturing the light source. However, it could be time-consuming, and it can complicate the real-time capturing, because it requires capturing, assembling, and aligning the range of images captured at different exposures. To overcome this obstacle, there was an option of using inverse tone-mapping algorithms for high-quality lighting that is based on linearizing and expanding the dynamic range of panoramic low dynamic range (LDR) images [60, 95, 96]. However, overexposed areas can be washed out resulting in information loss, thus making accurate reconstruction more challenging in large and overexposed areas.

Some studies indicated that the intensity of one bright light source in the scene could be estimated by placing a diffuse physical sphere in the scene and analyzing the captured image. While it is known that the mirrored sphere presents more accurate illumination, at least two images are required to estimate a single saturated light source [9, 19, 43].

Debevec et al. [19] used a novel design of light probe as a mirrored sphere divided into quadrants by diffuse strips. They demonstrated how to recover the full dynamic range of scene from a single exposure. The intensity of several saturated light sources could be estimated based on individually shot light probe image by a simple linear system solution.

A different approach is to set the intensity manually by extracting saturated regions of an LDR light probe. A more intuitive representation to work with is to convert the light probes into a sphere for the directional lights.

4.2 Temporally Variant Illumination (Video Sequences)

An accurate illumination and temporal consistency can also be accomplished in the synthetic object composited into real-world video sequences. The challenges of capturing the light probe at dynamic video frame rates exceed those of static scenes. Therefore, to address these challenges, multiple methods have been suggested to overcome the problem and produce acceptable results. Some successful methods used high dynamic range (HDR-video) cameras in real time attached to typical light probes and typical video cameras with distinctive optical filters or customized light probes [53, 60, 88, 95, 98, 105].

Knecht et al. [53] captured the surrounding environment and incident illumination by using a fish-eye camera where the system uses a pre-modeled real scene representation. They simulated some direct incident light with a small pocket lamp, then used the Studierstube Tracker to track the camera and lamp position. The default representation is 256 virtual point lights (VPLs) and 1,047 points per VPL for the scene. However, the results need some improvement to solve the wrong double shadowing and inconsistent coloring in some cases, and also a different shading method is used to increase the image quality. Kaufmann et al. [47] also captured the light in the real environment utilizing a fish-eye camera that allows the estimation of light source position and intensity.

Rohmer et al.'s [88] work involved extracting illumination information around the indoor environment by placing several HDR video cameras in the near-field environment. The extracted information was applied to the virtual object to simulate the current lighting condition where that object can change position freely with a consistent simulation of the illumination effects adapting to temporal deviations. The study took advantage of the distributed systems using a tablet camera and stationary PC where the source of light is not visible. Their work required external tracking and was designed for diffuse Lambertian objects, but it also provided a plausible display on glossy materials.

Unger et al. [105] formalized the transition of the old use of one or possibly multiple filtered frames of HDR panoramic to the use of HDR-video input. They adopted techniques from precomputed radiance transfer (PRT) to extend the dynamic processing and rendering of the real scene as input instead of a simulated video, see Figure 6.

Liu et al. [65] suggested tracking the illumination in the outdoor environment online using a complete image-based lighting method that varies from capturing a video with moving cameras. The intensity of the sun/skylight was captured by an optimized process to provide coherence in

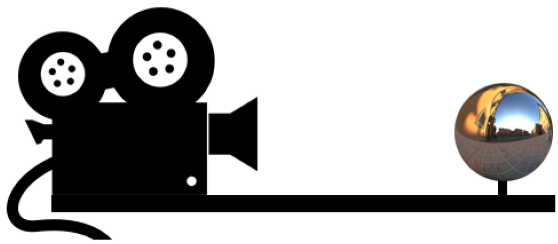

Fig. 6. Temporally variant image-based lighting example setup.

the temporal illumination. They used a set of real-life videos to demonstrate the visual coherence of the results along the video sequences. This study also covered the spatial variations and the estimation procedure.

Son et al. [97] obtained information about the light from the surrounding environment using the omnidirectional camera that has a typical low dynamic range (LDR) sensor to improve time-consuming in HDR. Furthermore, the light information can be transmitted to mobile devices to generate a realistic composition scene. Over the video sequence to generate a realistic result, the coherence between the spatial and temporal light should be maintained between the background and synthetic objects.

Gruber et al. [35] presented an approach that can handle the dynamic change of light sources and scene geometry with one portable RGB-D sensor in real time. Temporal consistency is an essential aspect of rendering with temporally varying illumination. A significant degree of visual noise is expected while depending on the light probe sequences as an input method. Temporal filtering or special rendering methods can fix the light probe sequence noise [11, 73].

Franke [26] introduced a relighting algorithm that computes the direct shadows and first bouncing indirect illumination called Delta Voxel Cone Tracing. The algorithm is temporally coherent and combined real and virtual surfaces with the extracted illumination at interactive rate.

4.3 Spatially Variant Illumination

There is a well-known limitation for the traditional image-based techniques regarding capturing the light when it varies based on the different locations in the real scene, i.e., illumination with spatially varying features. Shadowcasting and light shafting were factors in lighting designs that are used in visualization and cinematography production. There is a noticeable difference between the traditional rendering results of image-based lighting using omnidirectional light measurement, and rendering with a spatially varying scene illumination. Even though the result of conventional IBL rendering using a single HDR environment map looks realistic, it does not capture significant details about the scene lighting [104].

Spatially varying lighting environment $L(p', \vec{w}_i)$ measurement and representation requires capturing the scene lighting with angular distribution at several locations, and/or capturing the scene structure such as depth or parallax as a geometric model. These techniques are based on making assumptions and provide results in a different way. The amount of light measurements can vary greatly depend on using one HDR map or millions of samples. Also, the quality and accuracy of the reconstructed 3D geometry with details varies significantly by which computer vision algorithms was used, for instance, laser scanning, structure from motion (SfM), see Figure 7, or various pose estimation algorithms or others [10, 86, 104, 105].

Although using complex techniques could provide better accuracy, there is an obvious trade-off involved with these techniques such as processing time and user interaction. The amount of work and time spent by the user and the type of application are the factors that could determine the

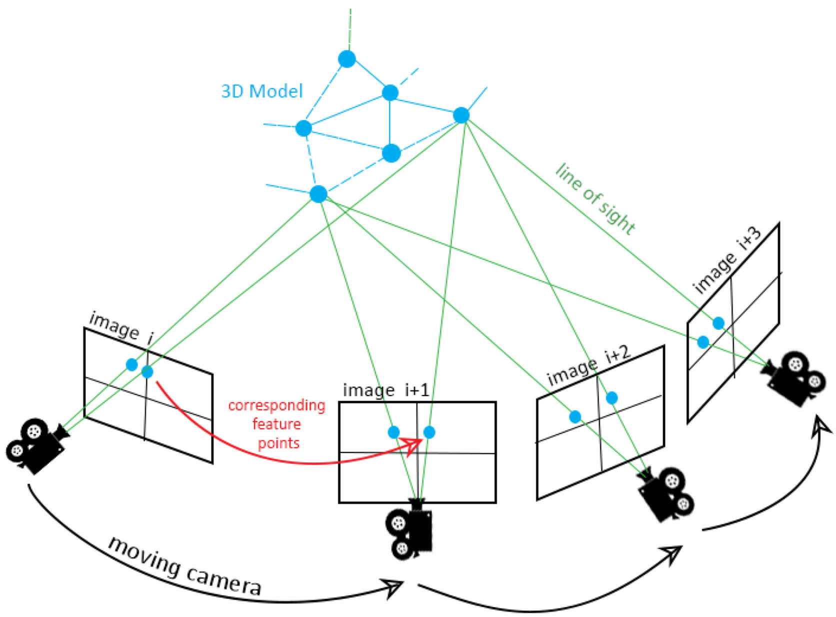

Fig. 7. An demonstrator of structure from motion (SfM) concept.

choice between the less complicated or more accurate methods. Most of the techniques produce plausible results. However, the methods that are less involved could result in a cruder illumination/lighting environment $L(p', \overrightarrow{w}_i)$ approximation. Therefore, the spatially varying illumination sampling [60] techniques could be divided into two main categories:

- Dense (heavy) point samples with no or a minimum number of geometric techniques that practice a significant amount of angular and spatial radiance samples to represent crude/no geometric scene model.
- Sparse (light) point samples with coarse geometry techniques assuming only Lambertian surfaces is the real-scene using a minimum number of omnidirectional HDR environment maps and representing a coarse geometric model for the scene.

Within each category, we will discuss some techniques and methods of spatially varying image-based lighting.

4.3.1 Dense (Heavy) Sampling. The techniques reviewed here are used for capturing and representing a subset, or part of the incident light field (ILF) of some synthetic objects placed at a specific region of the real scene during rendering. The ILF concept is related to the "light fields" notion in photography. The aim of this notion is capturing and processing the reflected/omitted outgoing light field as small sections of the real scene being captured as photographs. This concept facilitates the applications development of depth estimation, post-capture refocusing, and slight viewpoint transformations [40, 51, 60].

However, the ILF goal is totally the opposite to the light fields in capturing the light incident onto a certain region of the scene where interpolating nearby sampling points can estimate the full dynamic range for spatial and angular lighting variations $L(p, \overrightarrow{w}_i)$ [72, 88, 105].

Although most AR techniques are based on 3D and six degrees of freedom manipulation of real objects, there are times when 1D/2D procedures are required.

One-dimension spatial variation. Some of the old experiments used a set of dense HDR environment map images in the 1D path using an HDR-video camera from 1D light probe sequence to accurately capture and reproduce the spatial variation details in rendering. The difference between the rendering method used by Unger et al. [103] and the single HDR light probe environment lookup is the influence of the environment on the point of incidence instead of using the incident direction. The dataset enables the ability to capture light variations that cannot be handled using the traditional IBL, which is evidence that a light field of spatially variant illumination could produce a useful and powerful extension for the image-based lighting [103]. Most 1D algorithms in AR are used for tracking and interactions techniques. For more details, the reader is referred to [15, 107].

Nowrouzezahrai et al. [79] captured the environment lighting using a mirror sphere. They computed $L_{out}(p, \overrightarrow{w_o})$, which is the shade at a certain point x toward the view direction $\overrightarrow{w_o}$, which required the distribution of incident lighting at p, $L_{in}(p, \overrightarrow{w})$. The study assumed the spatial variation of lighting could be aggregated into the directional distribution that performed as an environment map of incident light where $L_{in}(p, \overrightarrow{w_o}) = L_{env}(\overrightarrow{w})$. The process of shading the virtual objects with real scene lighting increased the probability of consistent appearance among the virtual and real objects.

Two-dimensional spatial variation. Bradley et al.'s work [12] focused on aligning the virtual augmentations with a un-rigid plane or on flexible real objects such as cloth. They included an image-based approach for an automatic probing of the real-world light and shadow from HDR video. A 2D textured mesh rendering process aligned with the surface of the cloth and combined with the illumination result to achieve a non-rigid augmentation. The effect of diffuse reflection, ripples, and wrinkles can be seen in most of the cloth results where there is spatial variation in lighting. Therefore, if we replaced the diffuse light and shadow on the augmented objects, they could achieve a realistic output.

Frahm et al. [23] used fish-eye cameras and a TV camera for capturing the source of light in two images, because the image facing the camera usually escapes the light source. In these images, the light appeared as saturated regions, so the image sequence can be exploited for estimating the spatial samples of the light source. The image region is segmented automatically after the light sources segmentation. The mean assessment of the point samples belongs to the source of light, and the variance of these points are computed by the automatic segmentation. The light position in the reconstructed coordinate system is triangulated from different locations based on the light source. The reconstructions scale rotation of both cameras is subjugated to estimate the light source position that transformed into the front-camera coordinate system for more realistic augmentation.

Three-dimensional spatial variation. The concept of measuring and rendering the incident light field expanded to capture spatially varying illumination in 3D. Unger et al. [105] reconstructed a geometric scene model where the radiance information captured from the panoramic HDR video sequences with/without mirror sphere is re-projected and then stored as HDR texture to represent the direct sources of light and the surface of the scene.

Each image position and orientation is estimated using external tracking systems. For tracking purposes, they used a $1.5 \times 1.5 \times 1.5$ m translation stage along with an accuracy order of 0.1 mm and optical tracking based on tracking markers and external cameras. They computed two different datasets based on the image data. First, the dense point cloud that describes the scene geometry is estimated by structure from motion techniques on the HDR video frames. Second, the volumetric dataset of the illumination variations is computed from panoramic HDR video sequences that called focal volume. Each HDR-panorama pixel corresponds to a radiance sample that is picked from a particular position and direction in the scene.

The 3D dataset enables the user to improve the rapid estimation of the point cloud or model parts manually. The focal surface is placed inside the volume. The final geometry recovered from the scene reveals the robust light sources as high-intensity cones in the volume. The distinct voxels segmentation that corresponded to each light is enabled for position extracting and spatial extending using spatial selection and thresholding. Finally, the radiance samples are re-projected onto the recovered geometry that was stored as HDR textures. The recovered model during rendering of a point on an object is used to estimate the incident radiance.

4.3.2 Sparse (Light) Sampling. The complete capture of spatial variations of illumination measurements may not be practical in various situations. In rapidly changing environments like a film set, the changes should be captured prior to the next light scattering when the set re-organized. The techniques are used for this purpose to capture smaller samples of the HDR environment map such as one or two, only to enable high-speed and economical light capturing. Therefore, these methods are considered valuable in dynamic environments.

Computer vision and geometric relations are exploited by many of these methods to recover geometric scene information and provide spatial variation in the lighting. Lambertian surfaces are the default assumption of what the subjects are made up of in most studies. Even with the reasonable quality to capture the spatial variations effects, some of these assumptions are not as realistic as the other advanced methods that we will cover next.

One of the first methods was proposed by Sato et al. [91]; they used two omnidirectional cameras to generate the environment spatial radiance distribution utilizing stereo matching. The lighting information is presented as a 3D mesh that is used in the rendering process as an area of a light source that provides spatial variations in illumination.

Corsini et al. [17] suggested stereo light probes that used dual HDR environment maps instead of one. They used the spherical stereo to acquire not only the light sources direction and intensity but also the concrete position of these sources in the space [59].

Happa et al. [37] improved the lighting method in 3D modeled/scanned environment. Their approach relights synthetic interior scenes by IBL extension for generating great quality and fast interactive previews of the environments. It required light probes placed in the real scene and then manually aligned with the 3D mesh of the environment. Ultimately, the light is emitted from the HDR environment map has similarity to the instant radiosity or the photon mapping. Cook et al. [16] also present a method for the real-time photo-realistic rendering of photographs with synthetic objects in interior design by sampling each frame for diffuse shading to illuminate the virtual objects.

EnvyDepth developed by Banterle et al. [7] is a more general system that enables the user to paint or splatter depth onto an HDR environment map to generate a depth map using geometric constraints on primitives like planes, curves, and domes. From that understanding, this tool takes advantage of propagation editing for generating a detailed assemblage of virtual point lights that re-produce effects on both distant and local lighting in the real scene. The spatial information can provide more simulated effects such as shadow, highlights, and caustics compared to using single light in the distance. Therefore, without the struggle required for creating precise scene reconstructions, and without visible artifacts, EnvyDepth takes a few seconds only to produce plausible lighting.

A structured importance sampling is a technique introduced by Havran et al. [38], who were the first to use HDR video camera to render illuminated scenes by distant light based on Monte Carlo sampling.

5 REGISTRATION AND ESTIMATION OF LIGHTING METHODS

Registration is a transformation for the obtained sets of data that are the characteristic geometric and photometric properties, such as data obtained from sensors, day-time, depth maps, or perspectives, into a single coordinate system of the real/virtual scene. The process is required to compare and integrate the data acquired from this measurement.

The real scene physical measurements such as light probes and other devices are considered repetitive requirements and time-consuming. Therefore, images and videos are not acceptable as physical measurements in the real scene for this section. Many previous studies have concentrated on mining approximate lighting data directly from the images or video feeds to avoid these measurements in the real scene [60].

The direct lighting computation on a regular image is generally an ill-posed problem driving to the similar observation on that image with many potential solutions. Therefore, environment assumptions should be considered, for instance, known scene geometry, Lambertian reflectance, or observing illumination distribution [94].

The methods are categorized in this section according to the recovery procedure of the incident illumination from the original scene.

5.1 Explicit Geometric Registration

The methods used in this section are based on the reconstruction of scene geometry in a detailed manner that is usually recovered utilizing computer vision techniques, laser scanning, or manual modeling. The lighting data are estimated either using a few HDR images or several HDR environment maps, which could be represented as 2D texture mapping or surface light fields 4D function that is projected onto the scene geometric model [60].

It might be challenging to estimate and represent most of the spatial variations in scene illumination, like a sharp shadow, light shafts, and parallax effects accurately without estimating the 3D geometry of the scene. Nevertheless, many of these effects could be available if an accurate model has been recovered from the scene.

Debevec et al. [20] proposed a laser scanning system to capture an outdoor scene environment based on recovering geometry, capturing textures and lighting measurements. The 3D model projects the system textures and measurements of lighting to estimate the material properties at the scene surfaces using inverse global illumination techniques. The synthetic rendering output presented is taken from the photograph and by generated lighting setup.

Unger et al. [104, 105] presented a system for capturing, processing, and rendering virtual photo sets. The pipeline for capturing is image-based relying on SfM methods with dense geometry, and interactive tools set to estimate, and semi-automatically adjust the recovered scene geometry. The virtual photo set model composed of 3D geometry where the light information was captured from the projected HDR video sequence. The lighting data stored as a form of 2D textures or 4D surfaces light field. The article explored tools used to estimate the sources, position, and orientation of the light in the real scene, and the method to estimate the BRDF on surfaces with dense samples in the recovered model.

Meilland et al. [74] described another system that was involved with real-time 3D mapping and tracking using an RGB-D camera such as Kinect for recovering an irregular geometric scene. The observed dynamic range is employed for estimating the camera pose and the density of the scene structure, which fuse LDR exposure into light fields of HDR surfaces. The laser scanning is commonly used to recover scene geometry by surveying landmarks or hand modeling for visual effects productions. Recently, the development of captured/painted HDR texture tools has become more critical for realistic results [60, 105].

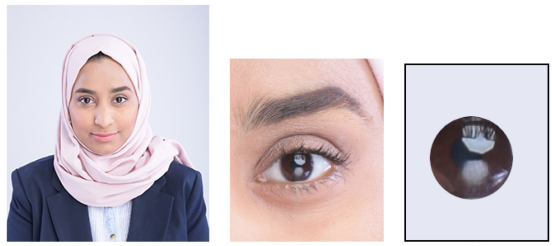

Fig. 8. (a) A frontal headshot. (b) A magnified eye cornea. (c) Using the eye pupil as an environment map.

5.2 Photometric Registration from Implicit Geometry

Familiar objects in several scenes with known or trivial geometry can be employed for estimating the incident illumination.

The notion of using the human eye as natural light probes was observed by Tsumura et al. [102] Then, Nishino et al. [78] took advantage of this observation to develop a robust framework to estimate the incident illumination by observing the eye cornea and acquiring the reflected scene radiance, see Figure 8.

Moreover, using the characteristic of the human face to estimate the real-world lighting condition was also exploited by Knorr and Kurz [54] to propose another framework. The model of face-appearance is explored from multiple faces dataset that was loaded offline under pre-known lighting conditions. The most acceptable lighting conditions of the real world in the bases of spherical harmonics are recovered at the runtime.

Some studies employed RGBD cameras, such as the Kinect sensor, to approximate and update the scene geometry dynamically. The low frequency of incident illumination can be recovered based on the Lambertian scene assumption that reflects temporal variations [10, 39, 73].

5.3 Photometric Registration from Specular Reflections

The incident light can be estimated from the reflected direction by observing the specular reflections on pre-registered objects. The concept can be used on any specular object in the scene with a recognizable shape without any constraints to any light probe.

For instance, Jachnik et al. [40] presented an approach to capture the incident light field from a certain surface by a particular camera with tracking browsing ability. They used surface light field instead of the more popular two plane light field parameterization. The surface of light field can be denoted as a 4D function $L(p, \overrightarrow{w_i}))$, where the planar case $p' = (p, d)$ is the texture element for 2D position in a simple Cartesian coordinate system. The radiance function depending on viewing direction was utilized instead of using a single RGB radiance value for texture element of particular surface. The function represented by discretizing the hemisphere surface and the value of a discrete point are stored. The hemisphere samples need to be spread evenly as much as possible for an efficient representation of the outgoing light distribution [40].

5.4 Photometric Registration from Diffuse Reflections

The attempt of computing the photometric registration from diffuse reflections seems viable when no specular reflections can be identified in the scene. The more common approach for indoor scenes is the diffuse surfaces. The more difficult type of inverse rendering problem is incident light recovery from diffuse surfaces, because separating the light contributions from different directions is crucial. Usually, the estimation only has a single dominant light direction.

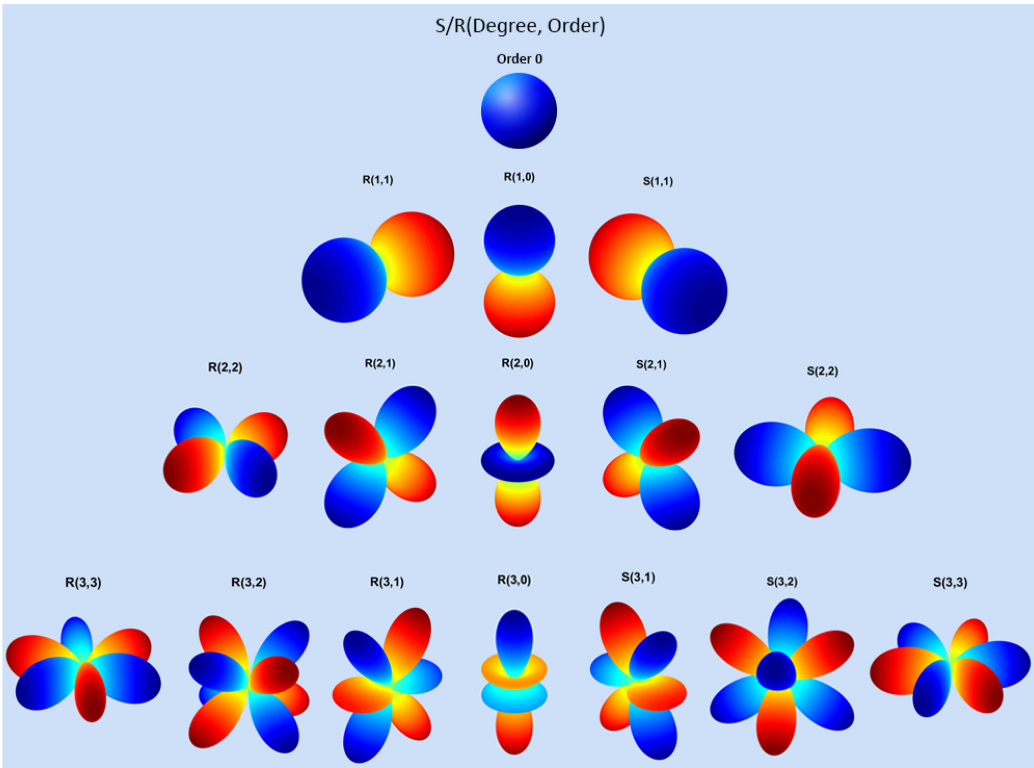

Fig. 9. The first four orders 0, 1, 2, 3 of the Spherical Harmonic function based on the origin and color distance on a sphere unit, Blue represents the positive values, and red represents the negative one.

Stauder [99] proposed the first video conferencing system capable of autonomously estimating an ambient light and single distant point light by estimating the scene geometry. The system obtained an estimate of directional light by estimating background segmentation of ellipsoid geometric models. Furthermore, Illumination can be stored in a more mathematical and consistent approach by using spherical harmonic (SH), which represents a 2D function on a sphere over all possible directions as the basis-functions set in a form of linear combination, see Figure 9 [84].

The low-frequency representation is usually sufficient for storage, because a few numerical coefficients per cache entry are needed to be compressed in SH form. Furthermore, The SH form to compute the diffuse light transport is very inexpensive and can be stored in a surface texture map [52].

Gruber et al. [34] demonstrated the SH framework ability for recovering lighting in the real scene in real time by employing only an RGB-D camera. The images of depth are used to reconstruct the scene. The directional incident light was solved by assuming only the SH form of diffuse reflection from sample points on the reconstructed surfaces. This distribution of surface normal should be represented efficiently with good sample points. While the diffuse reflection collected the light from every direction, a shadow from other objects should be computed for each sample point in the scene.

Also Gruber et al. [35] represented a desktop GPU system with optimizations of image-space that can estimate the light source and the shadow cast from a dynamic object with 20 fps. Also, Boom et al. [11] proposed a complete system for estimating one light source from the arbitrarily geometry of a scene assuming diffuse reflection for the whole scene. The method concept is an

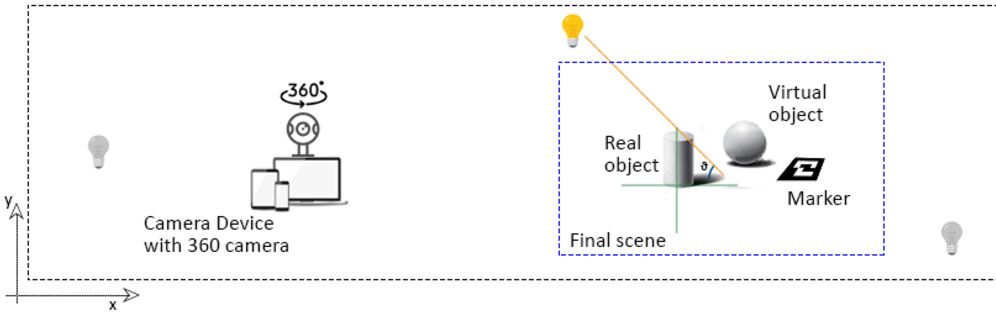

Fig. 10. Estimating the incident light direction by forming a ray from a unique point of the contour of the shadow that is a corresponding surface point on the shadow caster [5].

image segmentation based on color into super-pixels for assumed constant albedo also known as diffuse reflectance. The light source can be recovered with reasonable accuracy as long as the albedo is known.

5.5 Photometric Registration from Shadow

The light source can also be estimated using another method, which is by observing the shadow in an image, see Figure 10. The full or partial knowledge of the shadow caster's geometry, the correct classification, and the shadow appearance measurement in an image are the principles of this method.

The visible object geometry boundary of the shadow caster can be traced back by surface points on the contour. Therefore, estimating the direction of single or multiple light sources is applicable [94].

A complete overview of most known algorithms and methods based on shadow volume in computer graphics are discussed in detail by Kolivand and Sunar [55]. The algorithms were divided into two subsections: image based and geometry based.

Haller et al. [36] represented a special geometry characteristic employing a light probe called shadow caster that captured light from any direction. The more challenging approach is detecting shadows in natural images [10, 74, 96, 108].

As is well known, shadows are caused by the occlusion of incoming light in the real scene by objects. Therefore, to estimate the illumination distribution, the relation between the image brightness and the occlusion of incoming light can be analyzed. Sato et al. [92] introduced a technique to estimate the illumination distribution by investigating the image brightness inside shadow cast from a known shape object.

5.6 Photometric Registration from Images

The problem of inverse rendering was mentioned in detail by Patow et al. [81], where a curved object is photographed from various angles to capture the reflected light from a different orientation of the surface.

Marschner [70, 71] presents a system called image-based BRDF to measure reflectance quickly without any additional equipment that only required a series of photographs using a digital camera and stable light source.

Although Reconstructing real scene geometry from photographs using several viewpoints was implemented by Loscos et al. [67], a different method to recover reflectance was presented using controlled and fixed viewpoint to capture a set of images. A varying illumination without shadow and confidence weight factor are combined to represent the visibility under the light source consideration for every pixel in the images. The effect of shadow was considered later in Reference [66], which presents an algorithm for interactive relighting based on geometry reconstruction and not occluded illumination textures creation in a reprocessing step.

Poulin et al. [83] introduced a three-step process for 3D geometry reconstruction and texture extraction through an interactive system using a set of photographs. The authors solved a least-squares problem for the camera parameters first, then for the 3D geometry. The color textures of 3D model are extracted through re-projected texels sampling, when a satisfying model is retrieved. To form a unique texture, a fitting process is applied for all the textures then a combining process to the corresponding colors is achieved under certain criteria. Sato et al. [93] also used a sequence of images and a reconstructed 3D model to recover a simplified BRDF model know as Torrance-Sparrow reflection model for an isolated object.

However, Mandl et al. [69] proposed learning light probes that use pure synthetic images that are then applied on a real image dataset to train convolution neural networks (CNN) for high-quality illumination estimation. This method was developed to reduce the runtime in the inverse rendering techniques. Weber et al. [106] developed a deep learning method that was trained on a database of environment maps to estimate indoor lighting given a single image of a known 3D object.

Gardner et al. [27] trained a lighting classifier that is robust and annotates the light location by automatically using LDR environment map dataset. Then these annotations were used to train a deep neural network for light location and intensity prediction in a scene from a single and limited field of view image.

Morgand et al. [76] presented a geometric model that is reconstructed from images of specular material that depict the light sources on the planar surface to predict the shape of the specularities and the light source.

Grosch [32] combined inverse rendering for light reconstruction and differential rendering for displaying the changes to an interactive tool to maintain consistent lighting with real-time modification using panoramic images.

5.7 Photometric Registration for Outdoor Scenes

Typically, the scene's complete geometric model in outdoor AR systems is not accessible, which creates a more challenging photometric registration. However, during the daytime, the simple strongest light source is the direct illumination from the sun that drives the simple illumination model. The first estimate uses the analytical model of the sun. The light source of a large area such as the sky can be a secondary approximation. Shadow cues can be used in the image for enhanced outcomes [9, 64, 65, 98, 108].

Lalonde et al. [61] presented a method that calculates the probability distribution over the position of the sunlight and its visibility to estimate the scene illumination conditions using a single outdoor image from a large dataset of internet photographs.

Madsen et al. [68] assumed an outdoor scene to automatically detect the shadow in the image then determine the relation between the sky irradiance to the sun irradiance. The sun position was computed based on the date, time, and location on the Earth.

Kolivand et al. [56, 58] proposed a new and unique technique for realistically rendering the outdoor scenes in real time by taking into account the sky color with respect to the sun position that involves a shadow generating algorithm, Z-Partitioning: Gaussian and Fog Shadow Maps.

6 COMPOSITION AND GLOBAL ILLUMINATION METHODS

The final and single image resulting from merging the virtual objects into the real scene, called composition. In this part, we will cover the real-time global illumination methods that face two main dimensions of complexity.

The first dimension is the type of light transport that is being simulated. The simplest class of algorithms is a shadow that only allows the removal of light. Soft shadow and color bleeding are permitted by the diffuse global illumination where the light is strongly reflected off colored surfaces to nearby objects. The ideal addition is to apply well-known specular effects, for instance, reflection, refraction, and caustics, for selected objects. Thus, the highest complexity for the entire scene is allowed arbitrary diffuse and specular light transport.

The second dimension is concerned with the scene; all light transport can be pre-computed for static lighting in a static scene where the camera is the only moving part. Pre-computation could solve the online performance problems. However, it may require excessive computational storage and resources, particularly if the system supports specular effects. However, it is necessary, at least, to compute the effect of dynamic objects on the light transport for every frame in the scene. Thus, the highest computational cost in a scene is when it has dynamic objects and dynamic light transport [94].

The computational cost for the global illumination is determined by these two dimensions, in addition to the scene size. In large dynamic scenes accompanied with complicated light transport features, the real-time updates still depend on high-performance workstations and mobile devices.

Two rendering passes factorization are introduced by many modern global illumination methods for the more tractable approach. Several advantages are gained by this separation where the scene light transport is computed in the first pass, while the distributed light information that forms the final image is collected in the second pass.

More information about rendering is discussed in Section 7. Currently, this section is organized by the light transport complexity beginning with cast shadow on the scene, followed by diffuse global illumination to specular global illumination.

6.1 Shadows in Common Illumination

The 3D scene structure could be determined mentally by a human observer using the shadow cues. Sugano et al. [100] addressed the visual consistency where the lighting and shadow of virtual objects should match the real objects in AR scenes to increase the perceived realism. It is an expensive cost to compute reflection and shadow simultaneously; therefore, computing the shadow can be an acceptable alternative. The representation of a shadow can be computed in the first pass, while this representation could be used for the shading of the surface points of the final image in the second pass. The existent shadowing techniques are influenced by shadow volumes and shadow mapping [23, 41, 87, 98]. For objects enclosed in a frustum that lie in shadow with respect to a given light source and shadow-casting polygon, that frustum is known as shadow volume, and its sides are called shadow volume polygons [94].

Everitt and Kilgard [21] presented a shadow volume technique based on the standard feature of the GPU called the stencil buffer. The technique of shadow volumes has four passes: (1) Draw the scene without illumination (i.e., the scene in the shadow), (2) the front facing shadow volume polygons rasterization discretely increases the stencil buffer, (3) the back facing shadow volume polygons rasterization discretely reduces the stencil buffer, and (4) redraw the scene with every fragment that has zero value in the stencil buffer where the fragment is rendered illuminated and not in the shadow.

Other considerations that must be taken under common illumination include not only real object shadows or virtual object shadows separately but also a consistent shadow cast from real to virtual and the other way around. Haller et al. [36] developed the technique presented by Everitt and Kilgard to apply for the common illumination. In the first pass, render the shadows that are casting from virtual objects to the real ones. In the second pass, render every virtual object including received shadows from virtual or real objects.

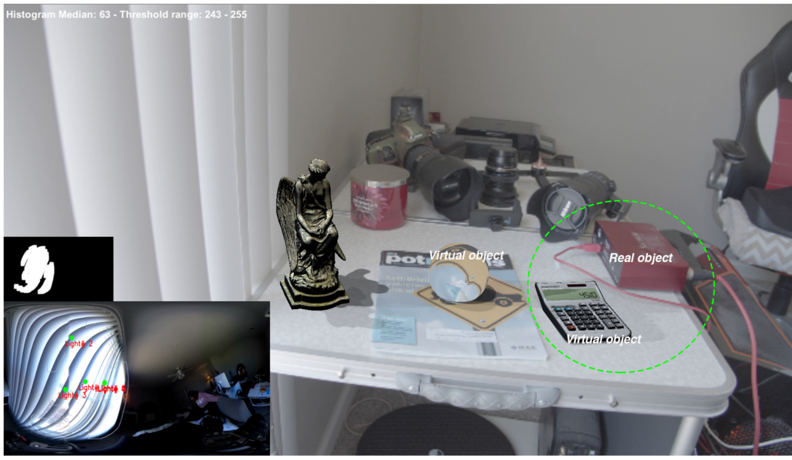

Fig. 11. Double Shadowing problem occurred when adding a virtual shadow of a virtual object (calculator) on a region in a shadow already from a real object (sound mixer) [3, 5, 6].

When the video feed initializes the frame-buffer, simultaneously the phantoms are rendered at first pass to the z-buffer, where the stencil buffer contains the shadow volumes of the virtual objects. The shadow cast from these objects to the real ones are created using the stencil buffer. All pixels of the video are blended to create an impression of the shadowed region by marking the stencil mask with a dark transparent color.

The second pass compares the traditional rendering of the stencil shadow volume. While the whole scene is rendered into the color buffer, the shadow volumes are drawn into the stencil buffer. A mask from the resulting stencil buffer then is used to redraw the whole scene with ambient and emissive components only, which makes the objects that fall into the shadow ray unlit by any light source available [2, 23, 94].

Shadow mapping technique is used in building the most contemporary rendering systems due to the fully accelerated texture mapping on the GPU. There are two passes for shadow mapping technique: In the first pass, the whole scene is rendered from the perspective of the light source to the depth buffer; in the second pass, the shadow map is used to determine whether a fragment is occluded also from the light source perspective to be rendered from the viewpoint of an observer. The fragment is considered part of the shadow if its depth in the coordinate of the light source is higher than the shadow map entry [23, 53, 73, 98].

Gibson et al. [28] and Supan et al. [101] and other old studies used shadow map to cast soft shadows from virtual to real objects by imposing shadow maps from a great number of light sources that were estimated using a light probe. In the current studies, the blending approach is applied to create soft shadows that can be pre-calculated even in the dynamic illumination scenes [73, 98].

6.2 Diffuse Global Illumination

Full global illumination requires not only perfect shadows but also a reflection. Some problematic issues can be eliminated using this approach. One of these problems is double shadowing where the shadow cast duplicates the darkening pixels from virtual to real objects if the blending is faulty due to virtual to real shadow overlapping with real to virtual shadow. Diffuse light transport of global illumination algorithms is the main focus of this section, see Figure 11.

The scene surfaces in the classic Radiosity method are turned into discrete small polygonal patches, and then the light transport among them is solved. The first pass is an inherently expensive procedure, because it requires global visibility computing among several patches. However, the second pass is a simple rendering process for the illuminated patches. Thus, in the dynamic scenes and the current real-time systems, patch-based Radiosity is hardly used. Fournier et al. [22] implemented the first application using the Radiosity method for simulating the common global illumination, but it was not in real time and performed with simplified assumptions.

Currently, most approaches aim for real-time performance at least in the second pass. A broad approach employs the shadow mapping on GPU for computing the direct illumination while simulating the indirect illumination with limited accuracy. For instance, Rohmer et al. [88] presented in their previous work how to store indirect illumination in an irradiance volume. For the first pass, radiance transfer from every possible direction onto the static scene is precomputed and combined into an SH form set of irradiance volumes. For the second pass, the shadow map is utilized to compute direct illumination, while the sum of basis irradiance volumes is weighted by the light intensity obtained in the indirect illumination.

Nowrouzezahrai et al. [79] also presented a light factorization algorithm for simple real geometry scenes. Separating the real-world lighting into two main lights, direct and indirect, is the main contribution of their work. For the direct light, they extracted the sources of point light from image-based lighting and employed the shadow mapping to apply them. The radiance transfer of each object is precomputed and represented in SH form that allowed an efficient combination of indirect light that is also represented in SH form [52].

Lensing and Broll [62] used a Virtual Point Lights (VPL) approach that is used in Knecht et al. [53] that we mentioned earlier, but they used splatting to apply the light form VPL instead of shadow mapping. Obtaining the dynamically moving geometry with deformable real objects using an RGB-D camera to support moving real objects was the critical contribution of the study. For computing illumination, the depth image with a guided edge-preserving filter used to smooth the massive noise in the depth image resulted in a better surface normal estimation. Beyond the current field of view, no real-world geometry existed, and every light source is virtual. They also presented a fast and novel global illumination method that depict indirect light for both diffuse and glossy surfaces for dynamic scenes in real time with sparse sampling using 3D geometry from a Kinect camera [63].

Franke [25] presented a global illumination technique that used a novel volumetric relighting method instead of surface light transfer. He utilized the light propagation volumes from other studies to represent radiance. The method computed a VPL set, and each VPL contribution is injected into a small volume with SH modeled directional radiance. The light propagation difference is computed before and after adding virtual objects to enable differential rendering.

Gruber et al. [35] estimated the light source by maximizing the prospect of global illumination interaction combining real-time data from three sources: the outside field of view (FOV) static geometric model, the inside FOV dynamic geometric model, and the lighting of environment dynamic estimation from reflections observed inside the field of view. Finally, the dynamic changes and occlusions that can be hard to observe outside the field of view were generally omitted. For more reading about the interactive occlusion in augmented reality see the research by Tuceyran et al. [13].

6.3 Specular Global Illumination

Specular effects from shiny surfaces, such as metal, and translucent materials, such as glass, are restricted in the view-independent methods described above.

Knecht et al. [53] demonstrated how to compute the specular effects in real time with their specular extension to Differential Instant Radiosity. Unfortunately, a rasterization approach cannot support arbitrary diffuse and specular combinations of light transport, which require a more expensive procedure based on ray-tracing.

Grosch et al. [31] introduced the first but not real-time method for specular global illumination. Ray-tracing for a photon mapping with a differential version is used in the first pass. Diffuse or specular is the classification of surfaces. The specular surfaces reflected or refracted the photons, but diffuse surfaces stored them. When a virtual object is hit by a photon, an anti-radiance, that is, a negative amount of light, is stored in the place where the photon would hit the real object. Then, ray-tracing from the eye is used in the second pass to produce a final image with reflections, refractions, and caustics affecting real imagery from virtual objects.

Kán and Kaufmann [47] used a related method based on the real-time ray-tracer OptiX proposed in Reference [80] for both passes then combined them with photon mapping. Moreover, instead of using anti-radiance, they enhanced the differential rendering in the second pass with separated shadow rays for virtual and real images.

Shi et al. [96] used the global illumination method to generate more realistic material appearances in both AR and synthetic material design. The differential irradiance caching algorithm presented in the previous article [47] was used combined with the ray-tracing that support enabling several bounces of the global illumination.

For evaluating the differential irradiance at the records of irradiance cache in one pass, they used Monte Carlo integration in GPU ray-tracing. The GPU accompanied with the NVIDIA OptiX ray-tracing engine provide parallel power to calculate both the direct and specular illumination [80].

7 DISPLAY AND RENDERING TECHNIQUES

Displaying your result is the final step to show the composited image to the user where the rendering methods are required. Most common rendering techniques could be divided into two categories:

- Common rendering algorithms for static or dynamic environment maps, which include Monte Carlo rendering, conversion to directional light sources, and pre-computed radiance transfer. More details of the material in this section are available in Reference [60].
- Interactive differential rendering methods, which are more involved with augmented reality and the illumination concept. Therefore, we will cover this concept in the following section.

Rendering a synthetic object into existing images has been developed over the years. Karsch et al. [50] proposed a method that creates a physical model of the scene that realistically inserts virtual objects into several photographs without any additional information about the scene measurements. The technique was suitable to render the objects with diffuse, specular, and glowing materials under the lighting conditions in the scene.

Gibson and Murta [29] also introduced first GPU-based AR system at real-time interactive rates using image-based method and sphere-mapping to eliminate and render synthetic objects seamlessly into background photographs.

For further evaluation of the existing rendering methods and techniques in computer graphics, Kolivand et al. [57] provided an elaborated survey on photorealistic rendering. The study aimed to ease the selection of the appropriate method in each system developed by researchers using classification and systematization among numerous methods.

7.1 Interactive Differential Rendering

The number of rendering passes that are required in the traditional differential rendering is two: one involved the local model of the real environment, and the other is merging both real and virtual objects. Many regions are rendered twice without any change, which raises a question about the visual effect under this approach. The use of a single pass is a more efficient approach where the changes in lighting created by the virtual objects are simulated directly.

According to real-world lighting conditions, we can compute the common illumination between any kind of objects using a real scene model, virtual scene, and incident light. The light traveling directly from the source to an object and reflected toward an observer is known as direct illumination. The light traveling from the source to an object and reflected toward another object is known as indirect illumination. A simulation of full global illumination can involve many light bounces between the objects before it eventually reaches the observer. The combinations of the object interactions could be any of these four possibilities: from real object to other real object or to a virtual object, and from virtual object to other similar virtual object or different real object. The composition of real and virtual is based on differential rendering that contributes to visual realism.

Common illumination even with a precise photometric registration would not be perfect, because it is not possible to fully interpret all light interactions in the scene. However, it would be efficient to preserve the subtle illumination effects that are present naturally in the real scene final image. The process of allowing the real-world illumination to be preserved is referred to as differential rendering. Fournier et al. [22] introduced the concept, then Debevec [18] developed the formula of differential rendering as follows.

Given the scene geometry, scene material, camera parameters, and light source, we can compute a light simulation L_R that corresponds to the original scene without virtual objects. A second light simulation $L_{(m)}$ can be computed after inserting the virtual objects. Any pixels depicting virtual objects can be replaced by $L_{(m)}$. For all pixels depicting real objects the difference $L_{(m)} - L_R$ shows the changes that happened to the real objects after adding the virtual objects. Then, the difference can be added as a correction term to the camera image L_c, see Figure 4.

Therefore, for pixels with virtual objects $L_{final} = L_{(m)}$, and for pixels with real objects $L_{final} = L_C + L_{(m)} - L_R$ can be interpreted as an error term to simulate the result $L_{(m)}$ for correction of any inaccuracies in the modeling L_R of the original scene L_C. The pixels are brightened if the virtual objects indirectly illuminate them $L_{(m)} - L_{(R)}(positive)$. However, the pixels are darkened if the virtual objects cast a shadow $L_{(m)} - L_{(R)}(negative)$.

This rendering could be more challenging if the scene modifications provide relighting, which changes the light sources and how that will affect the whole scene and not only the objects. Mainly, the idea is to remove the light source and cause the shadow to disappear from the scene. However, adding a new virtual light source would be applicable where the light can be linearly combined. Therefore, many methods were enhanced and developed to accommodate real-time global illumination.

Grosch et al. [31] modified photon mapping described in Jensen et al. [42] by using a differential photon mapping render in one pass for both real and virtual objects interaction. Every pixel of the environment map is representing a parallel light source. Thus, the photons are shot towards the virtual object: First, they are uniformly distributed on a disk that has a radius perpendicular to the light direction. Then, if the virtual object was hit by a photon, then the next intersection point is calculated using the real geometry and a negative flux is assign to that photon. Otherwise, if the virtual object does not intersect by a photon, it can be ignored, due to unchanged status to the light path.

Also, Grosch et al. [33] suggested a global illumination technique for indoor scenes in real time using diffuse materials by light probes. The representation of near-field reflected light in the room is updated by using the direct light from outside and a dynamic irradiance volume. The direct lighting also used sampling and shadow mapping.

Knecht et al. [53] presented methods that combined instant radiosity utilizing differential rendering that only needs one rendering pass for achieving the real-time performance in both diffuse and specular objects. Their method was extended for handling the reflective and refractive objects, and caustic effects by assuming the real objects geometry is static and given.

Kan et al. [47, 48] developed a method for interactive global illumination using photon mapping that allows caustic and reflective or refractive materials. Also, they developed a one-pass differential rendering method in real time by utilized irradiance caching. This irradiance separated real and virtual objects by analyzing both diverse ray types and intersection situations, which could be helpful in the computing process of differential irradiance. It is known that the real and differential irradiance are stored in the irradiance cache record that then can be utilized on the GPU for irradiance cache splatting. This method has some limitation regarding diffuse materials that required pre-computation stages. However, the results were reasonable for multiple bounce global illumination [49].

Lensing et al. [62] solved the pre-computation stage of a one-bounce diffuse indirect lightning using reflective shadow mapping. Also, to overcome the errors of the depth image, the method used was purely image-based with some guided filtering.

The development of differential rendering extended to mobile devices. Rohmer et al. [89, 90] reduced the computational cost for each light using tile-based rendering, in addition to frustum culling techniques tailored for AR systems and applications. Monroy et al. [75] presented a similar system that works in dynamic environments with the ability to scan the real scene and then projects onto a two-dimensional environment map that contains RGB+Depth data.

8 DISCUSSION AND COMPARISON

This section provides a summary and discussion about the whole review and the different methods for capturing and rendering in mixed reality scenes. We select some of these methods and compare them with others and observed each method's requirements, advantages, and drawback. The study of limitations is the key to improvement in the future work, and we need to highlight them for the next part.

It is very important to acknowledge that each method is not excluded from the others. These techniques could be included within each other or in a different part of the whole process to achieve the desired realism.

To explore each technique and make some comparison among these methods, some criteria were established to recognize the major difference and how that weighed in the final outcome. This major difference in aim, accuracy, and robustness between each method makes it challenging to conclude the compression.

Therefore, based on the information provided in each paper we attempt to restrict some of the useful information to write this section. We believe this indication will be helpful as a future reference for our work and others regarding usability, performance and more. Thus, the criteria are:

- **Required data.** the prerequisite data that are essential to be obtained at the beginning of the procedure or as part of sub-procedure in the whole system.
- **Assumptions.** some systems work assuming certain conditions such as the number of light sources, objects position, or the indoor/outdoor environment. These systems address their assumption at the beginning of the procedure, because it might not work perfectly if any of them were altered.
- **Outcome.** the solution provided by the system including the accuracy of the result under the previous criteria.

- **Drawback.** the limitations of the system based on the required data, assumptions, and outcomes. The future work usually starts by discussing and discovering these drawbacks and the attempt to solve them.

There could be more criteria to cover, but it requires insight and evaluation of each system that seems outside the scope of the current review. We might make an approximation for the time and effort for the rendering and processing, but the result would not be trustworthy without any quantitative data.

An overview of these criteria to compare the previous methods from selected papers are presented in Table 1. These papers were chosen based on two factors: (1) significance in their field, (2) provide an insight for our future work.

9 OPEN PROBLEMS AND FUTURE WORK

While interactive illumination for photorealistic scenes in augmented reality/mixed reality has been the major focus on many research and industrial issues for many years, still there are many open problems and future work that could be accomplished.

Most current methods are based on a static scene assumption that enables capturing spatial variation in scene illumination, which requires significant effort during capturing and processing. Also, developing techniques for estimating material properties on the surfaces in scenes could help find methods to measure and estimate varying spatial illumination. Computer vision improvement accompanied by enhanced capturing devices would highly develop this field. For instance, ARToolKit by DAQRI, Tango by Google [90], Vuforia by PTC Inc, and ARKit by Apple are examples of software development kits that use computer vision to provide more techniques for a robust inverse global illumination that could increase the accuracy of virtual objects simulation that affect the real scenes.

Furthermore, most of these kits are compatible with many development tools that could ease the work, such as Xcode for ISO developer adds these packages to existing IOS projects leveraging IOS rendering technologies, in Android Studio the Android developer can add the kit experiences to existing apps using Java APIs and C++ APIs. Also, it works with 3D engines like Unity and Unreal to build a game or high-fidelity 3D experience, then run this application on IOS or Android devices. Some also support developments of Universal Windows Platform (UWP) apps for selected Intel-based window devices, including Microsoft surfaces and HoloLens.

The generalization of existing methods for photorealistic objects added in images makes them limited. The most problematic area involves complex illumination environments and specular objects where future studies are an attempt to present specific assumptions to solve each problem and create a more robust and accurate physical estimation for illumination environments.

A similar problem that should also be subject to further research is in regard to the photorealistic object augmented in the video to provide a dynamic real-time illumination and features in mixed reality. The primary challenge is ensuring the consistency of estimating the material properties such as color and reflectance, with correct illumination, throughout the video feeds. Both the new and optimized schemes can be developed if the temporal coherency took into account by deriving it from the scene that could increase the challenge to allow different lighting conditions for spatial variation within the scene simultaneously. For example, we could save time and money for making a real-time online application for entertainment productions. Other limitations arise from the LDR content to composite photorealistic virtual objects into images or video. The availability of HDR would open many possibilities for improving more robust and general methods. The inverse tone mapping and recovering HDR lighting conditions from LDR media are some of the rousing methods for future researches.

Table 1. Comparison of the Overviewed Methods from the Selected Papers That We Covered Forcing on Four Criteria: Required Input, Assumption, Outcome, and Drawback

Method	Required Input	Assumptions	Outcome	Drawback
ACQUISITION OF LIGHTING				
Image-Based Lighting				
P. Debevec, 1998, [18]	HDR using a light probe.	Distance scene without reflectance model.	Acquire the real scene lighting effects with a good time and effort using differential rendering.	Limited physical accuracy there is no central projection point.
Schwandt et al. 2016 [95]	Single RGB-D camera LDR.	The light source is only a white light on the top of the scene.	Convincing reflections from the real scene and other virtual objects.	Incoherent reflectance on some surfaces that facing the camera.
Temporally Variant Illumination				
Knecht et al. 2010 [53]	LDR video using the fish-eye camera.	Available real scene geometric model with known martial properties.	Simulate the direct incident light and track the camera and lamp position.	Double shadowing and inconsistent coloring in some cases.
Liu et al. 2012 [65]	Set of real-life videos on moving the camera.	Access to information like daytime and GPS coordinates. Detect some planar surfaces.	Visual coherent of the results along with the video sequences.	Partial 3D data about the visible final scene is missing.
Son et al. 2012 [97]	Omnidirectional camera with LDR sensor.	Obtain only one light source.	Improve time-consuming in HDR.	Limited computation performance and bandwidth on mobile devices. No shadowing.
Gruber et al. 2015 [35]	RGB-D sensor.	Static geometry and illumination for the real scene, the light is white.	Handle the dynamic change of light sources and scene geometry with more 3D information.	Slow updating reconstruction, needs filtering to fix the noise. Only support the diffuse light.
Spatially Variant Illumination				
Unger et al. 2009 [72]	HDR video + light probe.	No major light field variation in the other two dimensions.	Capture and reproduce the spatial variation details in one dimension.	Limited spatial and angular resolution. Limited accuracy measurements.
Nowrouze-zahrai et al. 2011 [79]	HDR light probe.	The light spatial variation could be combined into a directional distribution. Static geometry.	Consistent appearance, solve light integration and dynamic visibility.	Shadow only works on static virtual objects. Difficulty in computing the soft shadow from the environmental light.
Unger et al. 2013 [104]	Multiple filtered frames of HDR panoramic.	Stationary scene.	Adopted techniques from pre-computed radiance transfer (PRT) to extend the dynamic processing and rendering.	Simultaneously capturing both the temporal and spatial domains.
Corsini et al. 2008 [17]	Stereo light probe (two reflective spheres).	The world coordinate origin is coincident with the spheres.	Estimate the light source positions robustly.	Acquiring light sources and scene geometry simultaneously.
Banterle et al. 2013 [7]	Single HDR light probe.	Distant lighting, relighting as photographs.	A plausible reconstruction of the local illumination environment.	The accuracy of modeling the scene based on the possible primitives. Occluded geometry.

(Continued)

Table 1. Continued

Method	Required Input	Assumptions	Outcome	Drawback
REGISTRATION AND ESTIMATION OF LIGHTING				
Explicit Geometric Registration				
Debevec et al. 2004 [20]	HDR light probes. Reflectance properties are reconstructed	Isotropic reflection surfaces where the light source must only move within a single plane of incidence.	Obtain a geometric model with illumination rendering consistent from real photographs and reflectance properties.	Surfaces with specular reflectance are not featured.
Meilland et al. 2013 [74]	LDR RGB-D.	The camera response function (CRF) is a simple non-linear with the auto shutter.	Dense 3D HDR environment-model estimation.	Complex to calibrate automatic shutter variations, only consider static objects, flickering.
Photometric Registration From Implicit Geometry				
Nishino et al. 2004 [78]	Environment map using the human eye as a probe.	The eccentricity and curvature radius at the apex, the same intensity for the entire image.	Estimate the illumination of the scene from imaging corneal system and relighting faces.	Limited dynamic range, the limited extent of reflections.
Knorr and Kurz 2014 [54]	A single image of a human face.	A distant light, close frontal head poses needed.	Estimate the real scene lighting condition in real time.	Only focus on the frontal pose, the approximation of radiance transfer function (RTF) is coarse.
Photometric Registration From A Specular				
Jachnik et al. 2012 [40]	Live Image of Specular surface.	Distant light, constant reflection, Consistent specular component through the whole surface, The surface radiance is proportional to the irradiance.	Dense illumination information in real time from the surface light field.	Limited dynamic range by the camera.
Photometric Registration From Diffuse Reflections				
Gruber et al. 2012 [34]	RGB-D sensor color Image and depth map.	Known scene geometry, diffuse surfaces (Lambertian reflectance model), the light color is white and distant.	Estimate light from the observed reconstructed model using SH.	Slow dynamic reconstruction, static camera, visual artifacts such as aliasing. The quality depends on surface normal vectors, only diffuse shadow, and lighting.
Gruber et al. 2015 [35]	RGB-D sensor color Image and depth map.	The light is distant and white. Estimate light with visual coherence in dynamic scenes.	Restrict to diffuse light transport and materials.	Geometry reconstruction quality and depth range are based on the sensor, visual artifacts.
Boom et al. 2017 [11]	RGB-D sensor color Image and depth map.	Diffuse surfaces (Lambertian reflectance model).	Estimate a point light position in the recorded scene.	Estimation of a Single point light source only, limited cast shadow.
Photometric Registration				
Haller et al. 2003 [36]	Image and depth information of the real objects.	The light does not fall directly on the non-rigid objects, distant scene.	Calculate the silhouettes to extract the shadow volume in real time and estimate the light source.	Shadow volume limited to a certain radius (distance between two objects), limited stencil volume size, and one light source only.

(Continued)

Table 1. Continued

Method	Required Input	Assumptions	Outcome	Drawback
Shi 2017 [96]	Two HDR video one for environment map and other for input.	Specular component not significant for the light source.	realistic reflection effects based on the physical lighting conditions It worked on limited parameter adjustments and BRDF models based on user edit.	Limited materials.
		Photometric Registration from Images		
Marschner 1999-1998 [70, 71]	A series of photographs and convexly curved samples with homogeneous BRDF	Stable light source	high resolution and accuracy of surface material on a large scale illumination and reflection directions without any special equipment.	Only work with flat samples.
Loscos et al. 1999-2000 [66, 67]	A series of photographs from different several viewpoint	Surfaces are diffuse.	Reconstructing real scene geometry and present a method to recover reflectance and interactive relighting considering shadow.	Does not include indirect lighting calculations.
Poulin et al. 1998 [83]	Set of photographs	Arbitrary camera parameters.	3D geometry reconstruction and texture extraction.	User interference, not an automated system.
Mandl et al. 2017 [69]	Pure synthetic images that then applied to real image dataset.	Known object in the scene	Train convolution neural network (CNN) for high-quality illumination estimation	Limited dataset, large CNN instances, not an automated system for object recognition.
Weber et al. 2018 [106]	single image, a database of environment maps	Known 3D object	Developing a deep learning method to estimate indoor lighting.	Must re-train different illumination prediction networks for other new objects and material properties. Must estimate geometry and material properties to train the neural network.
Gardner et al. 2017 [27]	a single and limited field of view image, LDR environment map dataset	Spherical Scene, ignore occlusion	A lighting classifier that is robust and annotate the light location automatically and train a deep neural network for light location and intensity prediction in a scene.	The warping operator cannot model occlusions. The method failed on images that have ambiguous photometric or geometric cues. The light size could be detected as smaller than it seems.
		Photometric Registration For Outdoor Scenes		
Lalonde et al. 2009 [61]	single outdoor image	The surfaces albedos are known, the sun visibility independent from its position.	Calculate the probability distribution over the position of the sunlight and its visibility to estimate the scene illumination conditions.	Some resulting estimations are weak. Several assumptions to reduce the problem complexity that could not be true all the time.
Liu et al. 2015 [64]	Single outdoor image.	The shadow casts on planar Lambertian surfaces, object position at world coordinate origin.	Estimate sunlight direction using the shadow cast on object modeling and recognition.	Limited range of the viewpoint, object bounding box with inaccuracy detection could lead to error in sunlight direction estimation.

(Continued)

Table 1. Continued

Method	Required Input	Assumptions	Outcome	Drawback
Barreira et al. 2018 [9]	GPS location, weather API for sky condition, ALS of the mobile device for illuminance measure.	Outdoor scenes.	Estimate the illumination condition then reconstruct the sun position and direction, detect dynamic shadow.	The object and mobile device should be in the same location. The object not halfway in shadow. Nonlinear color correction, the system only works in outdoor.
COMPOSITION AND GLOBAL ILLUMINATION				
Shadows In Common Illumination				
Ritschel 2011 [87]	High quality for Imperfect Shadow Maps (ISM).	Known geometric model of the scene and light information.	Compute interactive Indirect illumination in dynamic scenes Some spatial details could be lost.	The light should be smooth.
Everitt and Kilgard 2003 [21]	Triangles Models for occluding objects.	Ideal point light source. Detect the shadow area in the scene.	Incorrect shadow based on shadow depth count of the stencil value.	Ineffective use of the two passes render.
Diffuse Global Illumination				
Rohmer et al. 2015 [88]	Several HDR fish-eye video cameras.	The low frequency for the remaining illumination of the indirect radiance atlas.	Consistent simulation of the illumination effects adapting to temporal deviations.	Required external tracking and was designed for diffused objects more than other materials.
Franke 2013 [25]	Image and depth map from the RGB-D sensor.	Known scene geometry.	Visual coherence and relighting method for realistic indirect reflection between real and virtual objects.	Illumination bleeding and anti-radiance cause by some artifacts and the high cost of the procedure.
Specular Global Illumination				
Knecht et al. 2010 [53]	LDR video using the fish-eye camera.	Available geometric model of the real scene with known martial properties	Simulate the direct incident light and track the camera and lamp position.	Double shadowing and inconsistent coloring in some cases.
Kan and Kaufmann 2012 [47]	3D scene data, video image, environment image fish-eye.	Static objects.	High-quality specular effects with the depth of field effect and anti-aliasing.	Color bleeding and diffuse indirect lighting not featured, accurate camera tracking.

As we mentioned, the realistic real-time rendering that could capture and estimate the correct illumination is still an open area for research and development. It includes general interactions among real and virtual objects, glossy materials, dynamic scenes, and global illumination environments. Therefore, developing rendering algorithms for mixed reality especially for mobile devices is a significant part of the future work that will allow extensive embracing of photorealistic augmented reality. For instance, using photorealistic augmented reality systems for easing some challenges among mental health patients who have Parkinson's disease [8], Attention Deficit Hyperactivity Disorder (ADHD) [1], Autism Spectrum Disorder (ASD) [14], and phobia [44]. Improving the psychological presumptions of the human brain by making this application more realistic and well-blended with the real world could provide better results in the future.

Industrial organizations have been funding more academic augmented and mixed reality research. Many publications focused on topics that enhanced and served each structure's work requirements. The realistic results in global illumination would support the publicizing and advertising for sale of industrial products in general.

In the reverse process, these methods could be used by computer vision techniques for discovering forged or altered images, and the video and authenticity of these media could be used for forensic evidence, for instance. There are many applications and methods involved in this area, and by focusing on one problem at a time, many goals could be achieved.

ACKNOWLEDGMENTS

We thank every researcher and author of the original papers and publications for providing information and data that helped in the understanding of illumination and realism.

REFERENCES

[1] Nor Azlina Ab Aziz, Kamarulzaman Ab Aziz, Avijit Paul, Anuar Mohd Yusof, and Noor Shuhailie Mohamed Noor. 2012. Providing augmented reality based education for students with attention deficit hyperactive disorder via cloud computing: Its advantages. In *Proceedings of the 2012 14th International Conference on Advanced Communication Technology (ICACT'12)*. 577–581.

[2] Kusuma Agusanto, Li Li, Zhu Chuangui, and Ng Wan Sing. 2003. Photorealistic rendering for augmented reality using environment illumination. In *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003*. 208–216. DOI: https://doi.org/10.1109/ISMAR.2003.1240704

[3] A'aeshah Alhakamy and Mihran Tuceryan. 2019. AR360: Dynamic illumination for augmented reality with real-time interaction. In *Proceedings of the 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT'19)*. 170–174. DOI: https://doi.org/10.1109/INFOCT.2019.8710982

[4] A'aeshah Alhakamy and Mihran Tuceryan. 2019. CubeMap360: Interactive global illumination for augmented reality in dynamic environment. In *Proceedings of the 2019 IEEE Region 3 Technical, Professional, and Student Conference (SoutheastCon'19)*. IEEE.

[5] A'aeshah Alhakamy and Mihran Tuceryan. 2019. An empirical evaluation of the performance of real-time illumination approaches: Realistic scenes in augmented reality. In *Proceedings of the 6th Augmented Reality, Virtual Reality, and Computer Graphics (SALENTO AVR'19)*, Lucio Tommaso De Paolis and Patrick Bourdot (Eds.). Springer International, Cham, 179–195. DOI: https://doi.org/10.1007/978-3-030-25999-0_16

[6] A'aeshah Alhakamy and Mihran Tuceryan. 2019. Polarization-based illumination detection for coherent augmented reality scene rendering in dynamic environments. In *Advances in Computer Graphics through the 36th Computer Graphics International (CGI'19) in Cooperation with ACM SIGGRAPH and Eurographics*, Marina Gavrilova, Jian Chang, Nadia Magnenat Thalmann, Eckhard Hitzer, and Hiroshi Ishikawa (Eds.). Springer International, Cham, 3–14. DOI: https://doi.org/10.1007/978-3-030-22514-8_1

[7] Francesco Banterle, Marco Callieri, Matteo Dellepiane, Massimiliano Corsini, Fabio Pellacini, and Roberto Scopigno. 2013. EnvyDepth: An interface for recovering local natural illumination from environment maps. *Comput. Graph. Forum* 32, 7, 411–420. DOI: https://doi.org/10.1111/cgf.12249

[8] Yoram Baram. 2004. Closed-loop augmented reality apparatus. US Patent 6,734,834.

[9] João Barreira, Maximino Bessa, Luís Barbosa, and Luís Magalhães. 2018. A context-aware method for authentically simulating outdoors shadows for mobile augmented reality. *IEEE Trans. Vis. Comput. Graph.* 24, 3 (Mar. 2018), 1223–1231. DOI: https://doi.org/10.1109/TVCG.2017.2676777

[10] Jonathan T. Barron and Jitendra Malik. 2015. Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 8 (Aug. 2015), 1670–1687. DOI: https://doi.org/10.1109/TPAMI.2014.2377712

[11] Bastiaan J. Boom, Sergio Orts-Escolano, Xin X. Ning, Steven McDonagh, Peter Sandilands, and Robert B. Fisher. 2017. Interactive light source position estimation for augmented reality with an RGB-D camera. *Comput. Anim. Virt. Worlds* 28, 1 (2017), e1686. DOI: https://doi.org/10.1002/cav.1686 e1686 cav.1686.

[12] Derek Bradley, Gerhard Roth, and Prosenjit Bose. 2009. Augmented reality on cloth with realistic illumination. *Mach. Vis. Appl.* 20, 2 (2009), 85–92. DOI: https://doi.org/10.1007/s00138-007-0108-9

[13] David E. Breen, Ross T. Whitaker, Eric Rose, and Mihran Tuceryan. 1996. Interactive occlusion and automatic object placement for augmented reality. *Comput. Graph. Forum* 15, 3, 11–22. DOI: https://doi.org/10.1111/1467-8659.1530011

[14] Xavier Casas, Gerardo Herrera, Inmaculada Coma, and Marcos Fernández. 2012. A kinect-based augmented reality system for individuals with autism spectrum disorders. In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications - Volume 1: GRAPP, (VISIGRAPP'12)*. INSTICC, SciTePress, 440–446. DOI: https://doi.org/10.5220/0003844204400446

[15] Andrew I. Comport, Éric Marchand, and François Chaumette. 2003. A real-time tracker for markerless augmented reality. In *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003*. 36–45. DOI: https://doi.org/10.1109/ISMAR.2003.1240686

[16] Jon Cook, Simon Gibson, Toby Howard, and Roger Hubbold. 2003. Real-time photo-realistic augmented reality for interior design. In *Proceedings of the ACM SIGGRAPH 2003 Sketches & Applications (SIGGRAPH'03)*. Association for Computing Machinery, New York, NY, 1. DOI: https://doi.org/10.1145/965400.965427

[17] Massimiliano Corsini, Marco Callieri, and Paolo Cignoni. 2008. Stereo light probe. *Comput. Graph. Forum* 27, 2, 291–300. DOI: https://doi.org/10.1111/j.1467-8659.2008.01126.x

[18] Paul Debevec. 2008. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the ACM SIGGRAPH 2008 Classes (SIGGRAPH'08)*. Association for Computing Machinery, New York, NY. DOI: https://doi.org/10.1145/1401132.1401175

[19] Paul Debevec, Paul Graham, Jay Busch, and Mark Bolas. 2012. A single-shot light probe. In *Proceedings of the ACM SIGGRAPH 2012 Talks (SIGGRAPH'12)*. Association for Computing Machinery, New York, NY. DOI: https://doi.org/10.1145/2343045.2343058

[20] Paul Debevec, Chris Tchou, Andrew Gardner, Tim Hawkins, Charis Poullis, Jessi Stumpfel, Andrew Jones, Nathaniel Yun, Per Einarsson, Therese Lundgren, and Marcos Fajardo. 2004. *Estimating Surface Reflectance Properties of a Complex Scene Under Captured Natural Illumination.* USC ICT Technical Report ICT-TR-06.2004.

[21] Cass Everitt and Mark J. Kilgard. 2003. Practical and Robust Stenciled Shadow Volumes for Hardware-Accelerated Rendering. *CoRRcs.GR/0301002(2003).* http://arxiv.org/abs/cs/0301002.

[22] Alain Fournier, Atjeng S. Gunawan, and Chris Romanzin. 1993. Common illumination between real and computer generated scenes. In *Graphics Interface*. Canadian Information Processing Society, 254–254.

[23] Jan-Michael Frahm, Kevin Koeser, Daniel Grest, and Reinhard Koch. 2005. Markerless augmented reality with light source estimation for direct illumination. In *Proceedings of the Conference on Visual Media Production (CVMP'05)*. IET, 211–220.

[24] Banterle Francesco, Artusi Alessandro, Debattista Kurt, and Chalmers Alan. 2017. *Advanced High Dynamic Range Imaging* (2nd ed.). A K Peters/CRC Press. 328 pages. DOI: https://doi.org/10.1201/9781315119526

[25] Tobias Alexander Franke. 2013. Delta light propagation volumes for mixed reality. In *Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'13)*. 125–132. DOI: https://doi.org/10.1109/ISMAR.2013.6671772

[26] Tobias Alexander Franke. 2014. Delta voxel cone tracing. In *Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'14)*. 39–44. DOI: https://doi.org/10.1109/ISMAR.2014.6948407

[27] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.* 36, 6, Article 176 (Nov. 2017), 14 pages. DOI: https://doi.org/10.1145/3130800.3130891

[28] Simon Gibson, Jon Cook, Toby Howard, and Roger Hubbold. 2003. Rapid shadow generation in real-world lighting environments. In *Proceedings of the 14th Eurographics Workshop on Rendering (EGRW'03)*. Eurographics Association, Goslar, DEU, 219–229.

[29] Simon Gibson and Alan Murta. 2000. Interactive rendering with real-world illumination. In *Rendering Techniques 2000*, Bernard Péroche and Holly Rushmeier (Eds.). Springer, Vienna, 365–376. DOI: https://doi.org/10.1007/978-3-7091-6303-0_33

[30] Robin Green. 2003. Spherical harmonic lighting: The gritty details. In *Proceedings of the Archives of the Game Developers Conference*, Vol. 56. 4.

[31] Thorsten Grosch. 2005. Differential photon mapping-consistent augmentation of photographs with correction of all light paths. In *Proceedings of the Annual Conference of the European Association for Computer Graphics (Eurographics'05)*. 53–56.

[32] Thorsten Grosch. 2005. PanoAR: Interactive augmentation of omnidirectional images with consistent lighting. *Proceedings of the Annual Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications (Mirage'05)*. 25–34.

[33] Thorsten Grosch, Tobias Eble, and Stefan Mueller. 2007. Consistent interactive augmentation of live camera images with correct near-field illumination. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology (VRST'07)*. Association for Computing Machinery, New York, NY, 125–132. DOI: https://doi.org/10.1145/1315184.1315207

[34] Lukas Gruber, Thomas Richter-Trummer, and Dieter Schmalstieg. 2012. Real-time photometric registration from arbitrary geometry. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'12)*. 119–128. DOI: https://doi.org/10.1109/ISMAR.2012.6402548

[35] Lukas Gruber, Jonathan Ventura, and Dieter Schmalstieg. 2015. Image-space illumination for augmented reality in dynamic environments. In *Proceedings of the 2015 Conference on IEEE Virtual Reality (VR'15)*. 127–134. DOI: https://doi.org/10.1109/VR.2015.7223334

[36] Michael Haller, Stephan Drab, and Werner Hartmann. 2003. A real-time shadow approach for an augmented reality application using shadow volumes. In *Proceedings of the ACM Symposium on Virtual Reality Software and*

Technology (VRST'03). Association for Computing Machinery, New York, NY, 56–65. DOI: https://doi.org/10.1145/1008653.1008665

[37] Jassim Happa, Thomas Bashford-Rogers, Kurt Debattista, and Alan Chalmers. 2011. Rendering interior cultural heritage scenes using image-based shooting. In *Proceedings of the Annual Conference of the European Association for Computer Graphics (Eurographics'11).* 49–56.

[38] Vlastimil Havran, Miloslaw Smyk, Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. 2005. Importance sampling for video environment maps. In *Proceedings of the ACM SIGGRAPH 2005 Sketches (SIGGRAPH'05).* Association for Computing Machinery, New York, NY, 109–es. DOI: https://doi.org/10.1145/1187112.1187243

[39] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST'11).* Association for Computing Machinery, New York, NY, 559–568. DOI: https://doi.org/10.1145/2047196.2047270

[40] Jan Jachnik, Richard A. Newcombe, and Andrew J. Davison. 2012. Real-time surface light-field capture for augmentation of planar specular surfaces. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'12).* 91–97. DOI: https://doi.org/10.1109/ISMAR.2012.6402544

[41] Katrien Jacobs and Céline Loscos. 2006. Classification of illumination methods for mixed reality. *Comput. Graph. Forum* 25, 1, 29–51. DOI: https://doi.org/10.1111/j.1467-8659.2006.00816.x

[42] Henrik Wann Jensen and Niels Jørgen Christensen. 1995. Photon maps in bidirectional Monte Carlo ray tracing of complex objects. *Comput. Graph.* 19, 2 (1995), 215–224. DOI: https://doi.org/10.1016/0097-8493(94)00145-O

[43] Salma Jiddi, Philippe Robert, and Eric Marchand. 2016. Reflectance and illumination estimation for realistic augmentations of real scenes. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct'16).* IEEE, 244–249.

[44] M. Carmen Juan, Mariano Alcaniz, Carlos Monserrat, Cristina Botella, Rosa María Baños, and Belen Guerrero. 2005. Using augmented reality to treat phobias. *IEEE Comput. Graph. Appl.* 25, 6 (Nov. 2005), 31–37. DOI: https://doi.org/10.1109/MCG.2005.143

[45] James T. Kajiya. 1986. The rendering equation. In *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'86).* Association for Computing Machinery, New York, NY, 143–150. DOI: https://doi.org/10.1145/15922.15902

[46] Divya Kamboj and Wankui Liu. 2014. Improved variance cut algorithm for light source estimation in augmented reality. *Int. J. Comput. Appl.* 85, 19 (2014).

[47] Peter Kán and Hannes Kaufmann. 2012. High-quality reflections, refractions, and caustics in augmented reality and their contribution to visual coherence. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'12).* 99–108. DOI: https://doi.org/10.1109/ISMAR.2012.6402546

[48] Peter Kán and Hannes Kaufmann. 2013. Differential irradiance caching for fast high-quality light transport between virtual and real worlds. In *Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'13).* 133–141. DOI: https://doi.org/10.1109/ISMAR.2013.6671773

[49] Masayuki Kanbara and Naokazu Yokoya. 2004. Real-time estimation of light source environment for photorealistic augmented reality. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, Vol. 2. IEEE Cambridge, UK, 911–914. DOI: https://doi.org/10.1109/ICPR.2004.1334407

[50] Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem. 2011. Rendering synthetic objects into legacy photographs. In *Proceedings of the 2011 SIGGRAPH Asia Conference (SA'11).* Association for Computing Machinery, New York, NY. DOI: https://doi.org/10.1145/2024156.2024191

[51] Kevin Karsch, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, Hailin Jin, Rafael Fonte, Michael Sittig, and David Forsyth. 2014. Automatic scene inference for 3D object compositing. *ACM Trans. Graph.* 33, 3 (June 2014). DOI: https://doi.org/10.1145/2602146

[52] Jan Kautz, John Snyder, and Peter-Pike J. Sloan. 2002. Fast arbitrary BRDF shading for low-frequency lighting using spherical harmonics. *Render. Techn.* 2, 291–296 (2002), 1.

[53] Martin Knecht, Christoph Traxler, Oliver Mattausch, Oliver Mattausch, Purgathofer Werner, and Wimmer Michael. 2010. Differential instant radiosity for mixed reality. In *Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality.* 99–107. DOI: https://doi.org/10.1109/ISMAR.2010.5643556

[54] Sebastian B. Knorr and Daniel Kurz. 2014. Real-time illumination estimation from faces for coherent rendering. In *Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'14).* 113–122. DOI: https://doi.org/10.1109/ISMAR.2014.6948416

[55] Hoshang Kolivand and Mohd Shahrizal Sunar. 2013. Survey of shadow volume algorithms in computer graphics. *IETE Techn. Rev.* 30, 1 (2013), 38–46. DOI: https://doi.org/10.4103/0256-4602.107338

[56] Hoshang Kolivand and Mohd Shahrizal Sunar. 2014. Realistic real-time outdoor rendering in augmented reality. *PLoS ONE* 9, 9 (2014), e108334. DOI: https://doi.org/doi:10.1371/journal.pone.0108334

[57] Hoshang Kolivand, Mohd Shahrizal Sunar, Samira Y. Kakh, Riyadh Al-Rousan, and Ismahafezi Ismail. 2018. Photo-realistic rendering: A survey on evaluation. *Multimedia Tools Appl.* 77, 19 (2018), 25983–26008. DOI: https://doi.org/10.1007/s11042-018-5834-7

[58] Hoshang Kolivand, Mohd Shahrizal Sunar, Tanzila Saba, and Hatem Ali. 2019. ReLiShaft: Realistic real-time light shaft generation taking sky illumination into account. *Multimedia Tools Appl.* 78, 5 (2019), 6073–6092. DOI: https://doi.org/10.1007/s11042-018-6296-7

[59] Matthias Korn, Maik Stange, Andreas von Arb, Lisa Blum, Michael Kreil, Kathrin-Jennifer Kunze, Jens Anhenn, Timo Wallrath, and Thorsten Grosch. 2008. Interactive augmentation of live images using a HDR stereo camera. *J. Virt. Real. Broadcast.* 4, 12 (2008). DOI: https://doi.org/10.20385/1860-2037/4.2007.12

[60] Joel Kronander, Francesco Banterle, Andrew Gardner, Ehsan Miandji, and Jonas Unger. 2015. Photorealistic rendering of mixed reality scenes. *Comput. Graph. Forum* 34, 2, 643–665. DOI: https://doi.org/10.1111/cgf.12591 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12591

[61] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. 2009. Estimating natural illumination from a single outdoor image. In *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision.* 183–190. DOI: https://doi.org/10.1109/ICCV.2009.5459163

[62] Philipp Lensing and Wolfgang Broll. 2012. Instant indirect illumination for dynamic mixed reality scenes. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'12).* 109–118. DOI: https://doi.org/10.1109/ISMAR.2012.6402547

[63] Philipp Lensing and Wolfgang Broll. 2013. LightSkin: Real-time global illumination for virtual and mixed reality. In *Proceedings of the 5th Joint Virtual Reality Conference.* Eurographics Association, 17–24.

[64] Yang Liu, Theo Gevers, and Xueqing Li. 2015. Estimation of sunlight direction using 3D object models. *IEEE Trans. Image Process.* 24, 3 (Mar. 2015), 932–942. DOI: https://doi.org/10.1109/TIP.2014.2378032

[65] Yanli Liu and Xavier Granier. 2012. Online tracking of outdoor lighting variations for augmented reality with moving cameras. *IEEE Trans. Vis. Comput. Graph.* 18, 4 (Apr. 2012), 573–580. DOI: https://doi.org/10.1109/TVCG.2012.53

[66] Céline Loscos, George Drettakis, and Luc Robert. 2000. Interactive virtual relighting of real scenes. *IEEE Trans. Vis. Comput. Graph.* 6, 4 (Oct. 2000), 289–305. DOI: https://doi.org/10.1109/2945.895874

[67] Céline Loscos, Marie-Claude Frasson, George Drettakis, Bruce Walter, Xavier Granier, and Pierre Poulin. 1999. Interactive virtual relighting and remodeling of real scenes. In *Proceedings of the Conference on Rendering Techniques 1999*, Dani Lischinski and Greg Ward Larson (Eds.). Springer, Vienna, 329–340.

[68] Claus B. Madsen and Michael Nielsen. 2008. Towards probe-less augmented reality. *Position Paper, Computer Vision and Media Technology Lab, Aalborg University, Aalborg, Denmark* (2008).

[69] David Mandl, Kwang Moo Yi, Peter Mohr, Peter M. Roth, Pascal Fua, Vincent Lepetit, Dieter Schmalstieg, and Denis Kalkofen. 2017. Learning lightprobes for mixed reality illumination. In *Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'17).* 82–89. DOI: https://doi.org/10.1109/ISMAR.2017.25

[70] Stephen Robert Marschner and Donald P. Greenberg. 1998. *Inverse Rendering for Computer Graphics.* Citeseer.

[71] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. 1999. Image-based BRDF measurement including human skin. In *Proceedings of the Conference on Rendering Techniques 1999*, Dani Lischinski and Greg Ward Larson (Eds.). Springer, Vienna, 131–144.

[72] Vincent Masselus, Pieter Peers, Philip Dutré, and Yves D. Willems. 2003. Relighting with 4D incident light fields. *ACM Trans. Graph.* 22, 3 (2003), 613–620. DOI: https://doi.org/10.1145/1201775.882315

[73] Soham Uday Mehta, Kihwan Kim, Dawid Pajak, Kari Pulli, Jan Kautz, and Ravi Ramamoorthi. 2015. Filtering environment illumination for interactive physically-based rendering in mixed reality. In *Proceedings of the Eurographics Symposium on Rendering*, Vol. 1.

[74] Maxime Meilland, Christian Barat, and Andrew Comport. 2013. 3D high dynamic range dense visual SLAM and its application to real-time object re-lighting. In *Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'13).* 143–152. DOI: https://doi.org/10.1109/ISMAR.2013.6671774

[75] Rafael Monroy, Matis Hudon, and Aljosa Smolic. 2018. Dynamic environment mapping for augmented reality applications on mobile devices. In *Proceedings of the Conference on Vision, Modeling & Visualization (VMV'18).* 21–28. DOI: https://doi.org/10.2312/vmv.20181249

[76] Alexandre Morgand, Mohamed Tamaazousti, and Adrien Bartoli. 2018. A geometric model for specularity prediction on planar surfaces with multiple light sources. *IEEE Trans. Vis. Comput. Graph.* 24, 5 (May 2018), 1691–1704. DOI: https://doi.org/10.1109/TVCG.2017.2677445

[77] Brian Mount, Stephen Latta, Daniel McCulloch, Kevin Geisner, Jason Scott, Jonathan Steed, Arthur Tomlin, and Mark Mihelich. 2013. Augmented reality with realistic occlusion. US Patent App. 13/309,372.

[78] Ko Nishino and Shree K. Nayar. 2004. Eyes for relighting. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 704–711. DOI: https://doi.org/10.1145/1186562.1015783

[79] Derek Nowrouzezahrai, Stefan Geiger, Kenny Mitchell, Robert Sumner, Wojciech Jarosz, and Markus Gross. 2011. Light factorization for mixed-frequency shadows in augmented reality. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality.* 173–179. DOI: https://doi.org/10.1109/ISMAR.2011.6092384

[80] Steven G. Parker, James Bigler, Andreas Dietrich, Heiko Friedrich, Jared Hoberock, David Luebke, David McAllister, Morgan McGuire, Keith Morley, Austin Robison, et al. 2010. OptiX: A general purpose ray tracing engine. *ACM Trans. Graph.* 29, 4. DOI: https://doi.org/10.1145/1778765.1778803

[81] Gustavo Patow and Xavier Pueyo. 2003. A survey of inverse rendering problems. *Comput. Graph. Forum* 22, 4, 663–687. DOI: https://doi.org/10.1111/j.1467-8659.2003.00716.x

[82] Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory to Implementation.* Morgan Kaufmann.

[83] Pierre Poulin, Mathieu Ouimet, and Marie-Claude Frasson. 1998. Interactively modeling with photogrammetry. In *Proceedings of the Annual Confernence on Rendering Techniques 1998*, George Drettakis and Nelson Max (Eds.). Springer, Vienna, 93–104.

[84] Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'01).* Association for Computing Machinery, New York, NY, 497–500. DOI: https://doi.org/10.1145/383259.383317

[85] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting.* Morgan Kaufmann.

[86] Thomas Richter-Trummer, Denis Kalkofen, Jinwoo Park, and Dieter Schmalstieg. 2016. Instant mixed reality lighting from casual scanning. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'16).* 27–36. DOI: https://doi.org/10.1109/ISMAR.2016.18

[87] Tobias Ritschel, Elmar Eisemann, Inwoo Ha, James D. K. Kim, and Hans-Peter Seidel. 2011. Making imperfect shadow maps view-adaptive: High-quality global illumination in large dynamic scenes. *Comput. Graph. Forum* 30, 8 (2011), 2258–2269. DOI: https://doi.org/10.1111/j.1467-8659.2011.01998.x

[88] Kai Rohmer, Wolfgang Büschel, Raimund Dachselt, and Grosch Thorsten. 2015. Interactive near-field illumination for photorealistic augmented reality with varying materials on mobile devices. *IEEE Trans. Vis. Comput. Graph.* 21, 12 (Dec. 2015), 1349–1362. DOI: https://doi.org/10.1109/TVCG.2015.2450717

[89] Kai Rohmer and Thorsten Grosch. 2015. Tiled frustum culling for differential rendering on mobile devices. In *Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality.* 37–42. DOI: https://doi.org/10.1109/ISMAR.2015.13

[90] Kai Rohmer, Johannes Jendersie, and Thorsten Grosch. 2017. Natural environment illumination: Coherent interactive augmented reality for mobile and non-mobile devices. *IEEE Trans. Vis. Comput. Graph.* 23, 11 (Nov. 2017), 2474–2484. DOI: https://doi.org/10.1109/TVCG.2017.2734426

[91] Imari Sato, Yoichi Sato, and Katsushi Ikeuchi. 1999. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Trans. Vis. Comput. Graph.* 5, 1 (Jan. 1999), 1–12. DOI: https://doi.org/10.1109/2945.764865

[92] I. Sato, Y. Sato, and K. Ikeuchi. 2003. Illumination from shadows. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 3 (Mar. 2003), 290–300. DOI: https://doi.org/10.1109/TPAMI.2003.1182093

[93] Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuch. 1997. Object shape and reflectance modeling from observation. In *Proceedings of the 24thAnnual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97).* ACM Press/Addison-Wesley Publishing Co., USA, 379–387. DOI: https://doi.org/10.1145/258734.258885

[94] Dieter Schmalstieg and Tobias Höllerer. 2016. *Augmented Reality: Principles and Practice.* Addison-Wesley Professional.

[95] Tobias Schwandt and Wolfgang Broll. 2016. A single camera image based approach for glossy reflections in mixed reality applications. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'16).* 37–43. DOI: https://doi.org/10.1109/ISMAR.2016.12

[96] Weiqi Shi, Zeyu Wang, Metin Sezgin, Julie Dorsey, and Holly Rushmeier. 2017. Material design in augmented reality with In-Situ visual feedback. In *Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations (EGSR'17).* Eurographics Association, Goslar, DEU, 93–103. DOI: https://doi.org/10.2312/sre.20171198

[97] Wonyoung Son, Bodam Nam, Taehyub Kim, and Hyunki Hong. 2012. Using environment illumination information in mobile augmented reality. In *Proceedings of the 2012 IEEE International Conference on Consumer Electronics (ICCE'12).* 588–589. DOI: https://doi.org/10.1109/ICCE.2012.6161985

[98] Kevin Soulier, Matías Nicolás Selzer, and Martín Leonardo Larrea. 2016. Real-time estimation of illumination direction for augmented reality with low-cost sensors. In *Proceedings of the XXII Congreso Argentino de Ciencias de la Computación (CACIC'16).*

[99] Jürgen Stauder. 1999. Augmented reality with automatic illumination control incorporating ellipsoidal models. *IEEE Trans. Multimedia* 1, 2 (Jun. 1999), 136–143. DOI: https://doi.org/10.1109/6046.766735

[100] Natsuki Sugano, Hirokazu Kato, and Keihachiro Tachibana. 2003. The effects of shadow representation of virtual objects in augmented reality. In *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*. 76–83. DOI: https://doi.org/10.1109/ISMAR.2003.1240690

[101] Peter Supan, Ines Stuppacher, and Michael Haller. 2006. Image based shadowing in real-time augmented reality. *Int. J. Virt. Real.* 5, 3 (Jan. 2006), 1–7. DOI: https://doi.org/10.20870/IJVR.2006.5.3.2692

[102] Norimichi Tsumura, Minh Nguyet Dang, Takao Makino, and Yoichi Miyake. 2003. Estimating the directions to light sources using images of eye for reconstructing 3D human face. In *Proceedings of the Color and Imaging Conference* 2003, 1, 77–81.

[103] Jonas Unger, Stefan Gustavson, and Anders Ynnerman. 2007. Spatially varying image based lighting by light probe sequences. *Vis. Comput.* 23, 7 (2007), 453–465. DOI: https://doi.org/10.1007/s00371-007-0127-6

[104] Jonas Unger, Joel Kronander, Per Larsson, Stefan Gustavson, Joakim Löw, and Anders Ynnerman. 2013. Spatially varying image based lighting using HDR-video. *Comput. Graph.* 37, 7 (2013), 923–934.

[105] Jonas Unger, Joel Kronander, Per Larsson, Stefan Gustavson, and Anders Ynnerman. 2013. Temporally and spatially varying image based lighting using HDR-video. In *Proceedings of the 21st European Signal Processing Conference (EUSIPCO'13)*. 1–5.

[106] Henrique Weber, Donald Prévost, and Jean-François Lalonde. 2018. Learning to estimate indoor lighting from 3D objects. In *Proceedings of the 2018 International Conference on 3D Vision (3DV'18)*. 199–207. DOI: https://doi.org/10.1109/3DV.2018.00032

[107] Harald Wuest, Florent Vial, and Didier Stricker. 2005. Adaptive line tracking with multiple hypotheses for augmented reality. In *Proceedings of the 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. 62–69. DOI: https://doi.org/10.1109/ISMAR.2005.8

[108] Guanyu Xing, Yanli Liu, Xueying Qin, and Qunsheng Peng. 2011. On-line illumination estimation of outdoor scenes based on area selection for augmented reality. In *Proceedings of the 2011 12th International Conference on Computer-Aided Design and Computer Graphics*. 439–442. DOI: https://doi.org/10.1109/CAD/Graphics.2011.51

Received September 2018; revised July 2019; accepted March 2020