Photometric Registration by Adaptive High Dynamic Range Image Generation for Augmented Reality

Yusaku Nishina Bunyo Okumura * Masayuki Kanbara * Naokazu Yokoya *
Nara Institute of Science and Technology (NAIST)

Abstract

This paper describes photometric registration for augmented reality (AR) using a high-dynamic-range (HDR) image. In photorealistic AR, estimating the lighting environment of virtual objects is difficult because of low dynamic range cameras. In order to overcome this problem, we propose a method that estimates the lighting environment from an HDR image and renders virtual objects using an HDR environment map. Virtual objects are overlaid in real-time by adjusting the dynamic range of the rendered image with tone mapping according to the exposure time of the camera. The HDR image is generated from multiple images captured with various exposure times. We have found through experimentation that the updating rate is improved by effectively limiting the dynamic range, depending on the exposure time. We have verified the effect of limiting the dynamic range on the reality of virtual objects.

Index Terms: H.5.1 [Information interfaces and presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities;

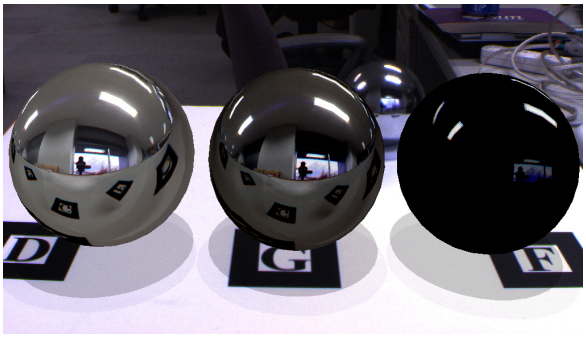
1 Introduction

In augmented reality (AR), we need to resolve the problems of geometric and photometric registration between the real and virtual worlds in real-time. Geometric registration means that a virtual object is located at an appropriate place in the real world , while photometric registration means that a virtual object is represented consistently in the real world in terms of shading, shadow, and image quality. Several studies have focused on photometric registration in photorealistic AR, in which a synthesized scene should appear realistic [8].

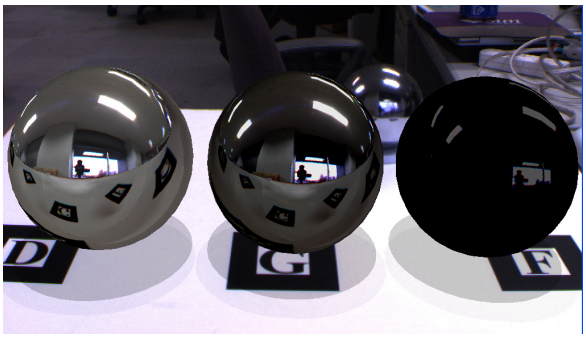
In order to appropriately represent shadow and shading of a virtual object in AR, it is necessary to estimate the lighting environment, i.e., the distribution of radiance in the real environment. For rendering the shadow of a virtual object, we need to estimate a light source with high radiance, called a direct light source. Representing the surface of a virtual object reflecting real world conditions involves acquisition with not only direct light sources but also indirect light sources, such as a normal object. In particular, for photorealistic AR, radiances of all objects including both a direct light source and a normal object in the real environment must be estimated.

Figure 1 illustrates examples of a photometric AR application. In the images, virtual objects with different albedo values, which is a reflectance parameter, are rendered by environment mapping with and without a high dynamic range (HDR) image. In Figure 1(a), the reflectance on the surface of virtual objects is correctly expressed because an outdoor scene can be appropriately rendered in the window region. However, since an HDR image cannot be precisely estimated due to the insufficient dynamic range, the window region in Figure 1(b) is not rendered correctly.

*E-mail: {bunyo-o, kanbara, yokoya}@is.naist.jp



(a) Synthesized image with HDR rendering.



(b)Synthesized image without HDR rendering.

Figure 1: Examples of photorealistic AR: Virtual spheres with different albedo values are rendered by environment mapping.

Related work on photometric registration for AR is classified into two major approaches. In the first, only direct light sources are used for the estimation of the lighting environment, e.g., estimation of the position of the light source by analyzing the shadow of a real object whose shape is known. The other approach considers both direct and indirect light sources for estimation, e.g., direct observation of the lighting environment with a camera and capturing the reflectance of a mirrored surface [2]. These methods attempted to acquire not only direct but also indirect light sources.

In order to overcome the problem caused by insufficient dynamic range of the camera, many approaches to acquire an HDR image have been proposed [4, 11]. These methods follow two major approaches using a standard camera whose dynamic range is not high. In one approach, the HDR image is generated from multiple images captured by a camera with varying exposure times [9]. However, the relationship between the accuracy of an HDR image and the generation time is a trade-off because the HDR image can be generated using many images captured with different exposure times.

In the other approach, an HDR image is obtained by adjusting the amount of light incident on each pixel with a filter attached in front of the camera [1]. This approach has an advantage that the frame rate is similar to the video rate. However, a special camera is needed and the dynamic range of the acquired image is still limited.

In a computer graphics, to represent a virtual object in a real world by photometric effects, an HDR rendering method with sufficient dynamic range in the rendering process has been investigated

[3]. Debevec proposed an HDR rendering method with environment mapping using HDR images [7, 6]. For displaying the images rendered using HDR images, the dynamic range of the images is needed to be compressed into the range of the display device. A tone mapping technique that appropriately maps a set of colors to the display device was proposed [5].

A method of photorealistic image generation using HDR images generated from multiple images captured with different exposure times was proposed for a mixed reality application that does not need real-time processing [7]. In a related study which uses HDR images for estimation of lighting environment in AR by assuming that the environment is static, Agusanto [2] realized photorealistic AR using HDR images captured in advance. However, since an AR environment is typically dynamic, HDR images need to be generated in real time. As mentioned above, there are two major approaches to generate HDR images: one is to use multiple images captured with different exposure times and the other is to use special equipment such as a filter attached in front of a camera. We adopt the former because the dynamic range is less limited. However, in this approach, the frame rate decreases as the width of the dynamic range increases.

This study is conducted to achieve photorealistic AR that can represent a realistic reflectance of a virtual object under a dynamic environment. In order to overcome the limitation of the dynamic range of light estimation, we adopt an approach involving HDR image generation using multiple images with different exposure times. By assuming that the dynamic range representing shading, shadow, and reflectance of virtual objects in photorealistic AR can be limited, we speed up the process of HDR image generation by adjusting the dynamic range needed for rendering virtual objects. The reflectance of virtual objects can be realized by HDR rendering with environment mapping of HDR images acquired in the real world. The shading and shadow of virtual objects are rendered by estimating the direction of the light source with respect to the HDR images. The proposed method realizes photorealistic AR that can render virtual objects adjusted to the lighting environment in the real world. The frame rate of HDR images can be increased by adjusting the dynamic range needed for rendering the virtual objects.

2 PHOTOREALISTIC AR USING HDR IMAGE

2.1 Overview

This paper proposes a video see-through AR system with two cameras, which capture the view from the user's viewpoint and the lighting environment. Figure 2 illustrates a flowchart of the proposed method. In this method, two processes, AR synthesized image generation and estimation of lighting environment, run asynchronously and in parallel.

A. Lighting environment estimation

To estimate the lighting environment with HDR, multiple images are captured with different exposure times (A-l). The HDR image is generated from the intensity of pixels and exposure time of the images (A-2), and the direction of the light source is estimated using the HDR image in order to represent the shadow and shading of the virtual object (A-3). Next, the range of radiance needed to render shadow, shading, and reflectance of the virtual object in the next frame is estimated (A-4). Finally, the exposure times and the number of images that can measure the range of radiance required for the next frame are determined (A-5).

B. AR image generation

First, a user's view image is captured by a camera attached to the user's viewpoint (B-l). Next, geometric registration is accomplished by estimating the relationship between the marker and camera coordinate systems by resolving a Perspective-n-Point (PnP) problem with detected corners of markers whose shape and color are known (B-2). When the virtual object is rendered using the position and orientation of the user's view camera, the shading and shadow of the virtual object are presented by computing the direction of light sources, and the reflectance of the virtual object is rendered with environment mapping using the HDR images generated in the process of lighting environment estimation [10](B-3). Finally, in order to realize the consistency between the dynamic ranges of the user's view image and the virtual object rendered with the HDR image, the virtual object is synthesized on the user's view image with a tone mapping technique (B-4).

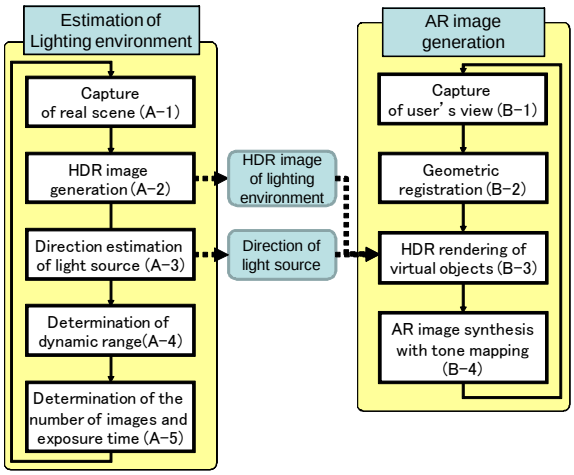


Figure 2: Flow diagram of the proposed method.

The processes of AR image generation and lighting environment estimation run in parallel and asynchronously, and share information of generated HDR images and directions of light sources. Section 2.2 and 2.3 describe the determination processes of dynamic range needed in the next frame (A-4) and the number of image and exposure times (A-5), respectively. It should be noted that main contribution of the paper is in an adaptive selection of exposure times for acquiring HDR image.

2.2 Determination of dynamic range needed in the next frame (A-4)

When an HDR image is correctly generated from multiple images, the maximum intensity E_{envmax} and the minimum intensity E_{envmin} of the HDR image are the maximum and minimum radiances in the real environment, respectively. The dynamic range of the scene is defined by the ratio of E_{envmax} to E_{envmin}. To reduce the calculation cost, the dynamic range required for the HDR image generation is adaptively changed to the dynamic range needed to represent the virtual objects in each frame. The maximum radiance and the minimum radiance to be required are decided by the exposure time of the user's view camera, albedo of the virtual object, and the maximum and minimum radiances in the real environment.

Figure 3 illustrates radiances for representing a reflectance of virtual object when a maximum and minimum albedo values are S_{max} and S_{min} $(0 < S_{min} < S_{max} \leq 1)$. E_{armin} denotes the minimum radiance required to represent a virtual object whose albedo is S_{max}, while E_{armax} meens the maximum radiance required to represent a virtual object whose albedo is S_{min}.

Provided that the exposure time of the user's view camera is ΔT_{ar}, the maximum radiances E_{armax} and the minimum radiance E_{armin} needed to be measured by the camera for acquiring the lighting environment are defined by Eqs. (1) and (2), respectively.

$$E_{armin} = \frac{\exp(G_{ar}(Z_{black}) - \ln \Delta T_{ar})}{S_{max}}, \quad (1)$$

$$E_{armax} = \frac{\exp(G_{ar}(Z_{white}) - \ln \Delta T_{ar})}{S_{min}}, \quad (2)$$

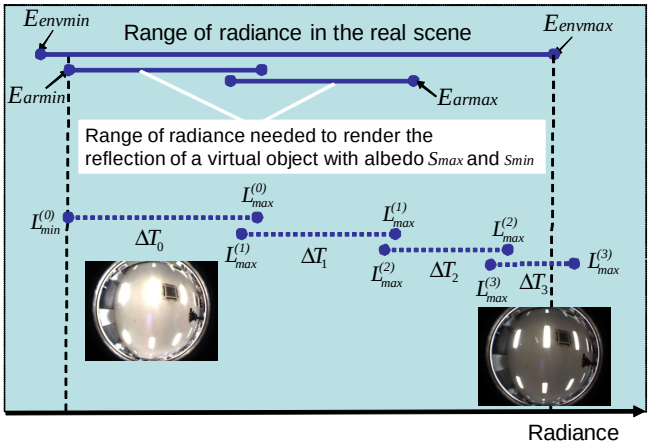


Figure 3: Range of radiance needed to render the reflection of a virtual object.

where G_{ar} is a response function of the intensity of a pixel and irradiance, and Z_{white} and Z_{black} represent intensities of pixels that are inappropriate white and black parts in the image. S_{max} and S_{min} indicate the maximum and minimum albedo values of the virtual object. Since the shape and color of the shadow of a virtual object are affected by a part with high radiance such as a light source, the maximum radiance E_{envmax} in the real environment must be estimated. If the HDR image is accurately generated and the change in dynamic range in each frame is small, we assume the maximum radiance E_{envmax} to be the maximum intensity of the HDR image in the previous frame.

The maximum radiance that is needed to be measured is the radiance E_{envmax}, in order to represent the shadow and shading of the virtual object. On the other hand, the minimum radiance needed to be measured is decided by the radiance E_{armin} needed for representing the reflectance of the virtual object. However, if the minimum radiance E_{envmin} in the real environment is higher than E_{armin}, it is not necessary to measure a radiance lower than E_{envmin}. To render a virtual object with an HDR image with sufficient dynamic range as shown in Figure 3, the radiance E to be measured obeys the relation in Eq. (5).

$$max(E_{envmin}, E_{armin}) \leq E \leq E_{envmax}. \quad (3)$$

2.3 Determination of the number of images and exposure times (A-5)

This section explains a method for determining the exposure times and the number of images based on the necessary radiance determined in Section 2.2. Note that the capture time includes the sum of exposure times, time taken for changing the exposure time, and time spent for transfer of the captured images. The capture time decreases if the number of captured images decreases. Since the capture time for the least number of images and the shortest sum of exposure times is required, our approach first determines the exposure time for minimum radiance.

The measurement accuracy of radiance on a pixel decreases as the pixel approaches inappropriate black or white parts. The maximum irradiance $L_{max}^{(j)}$ and minimum irradiance $L_{min}^{(j)}$ that can be estimated from an image are defined by intensities Z_{black} and Z_{white}, which are assumed to be the intensities of the improper parts. The irradiances $L_{max}^{(j)}$ and $L_{min}^{(j)}$ of the image captured in exposure time ΔT_j are given as follows:

$$L_{min}^{(j)} = \exp(G_{env}(Z_{black}) - \ln \Delta T_j), \quad (4)$$

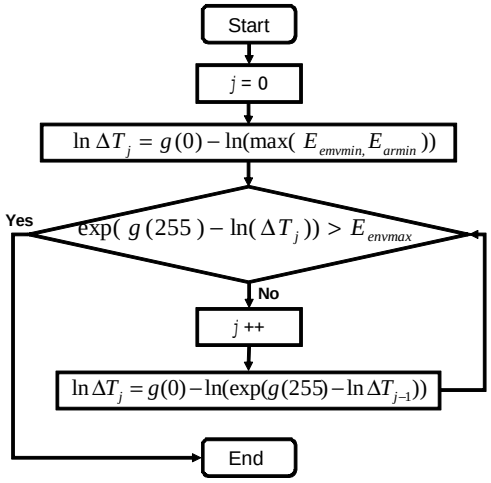


Figure 4: Flowchart for determining the number of images and exposure times.

$$L_{max}^{(j)} = \exp(G_{env}(Z_{white}) - \ln \Delta T_j). \quad (5)$$

The number of images and exposure time are decided by using the below conditions:

- The entire range estimated in Section 2.2 is covered.
- The number of images is the least.
- The sum of exposure times is the shortest.

To satisfy these conditions, minimum radiance $max(E_{envmin}, E_{armin})$ in the range to be acquired is first measured, and the next higher radiance is measured in order. Finally, if the maximum radiance E_{envmax} can be acquired, the number of images and exposure time can be determined. Figure 4 illustrates the flowchart for determining the number of images and exposure times. As shown in Figure 3, first, let the number j of the image be 0, ΔT_j is decided when the irradiance $L_{min}^{(j)}$ is equal to $max(E_{envmin}, E_{armin})$. Next, an image is captured with the exposure time which irradiance $L_{min}^{(j)}$ is $L_{max}^{(j-1)}$ in sequence until $L_{max}^{(j)} > E_{envmax}$. By this procedure, the appropriate number of images and exposure times that can acquire the range estimated in Section 2.2 can be determined.

3 Experiments

3.1 Experimental environment

To confirm the validation of the proposed method, we performed two experiments. First, we show the change in exposure time and the number of captured images by varying the real environment. In addition, the images synthesized in the real scene are also shown. Experiments are performed using a desktop PC (CPU: Pentium 4 3.4 GHz; memory: 2.0 GB; videocard: GeForce 6800GT; video memory: 256 MB), an IEEE1394 camera for user's view (Pointgrey Flea, resolution: 1024×768 pixels; frame capture rate: 30 fps), and a camera with a wide conversion lens for lighting environment estimation (Pointgrey Dragonfly Express, resolution: 640×480 pixels, maximum frame rate: 120 fps).

3.2 Evaluation of the number of images and exposure times

We show the estimation result when the albedo value of the virtual object is changed. First, we render a virtual object with an albedo of 1, which means the surface is like a mirror. After this, we render a virtual object with an albedo of 0.1. Note that the exposure time of the user's view camera and the real scene are constant. Figure 5 illustrates the required dynamic range and exposure times of

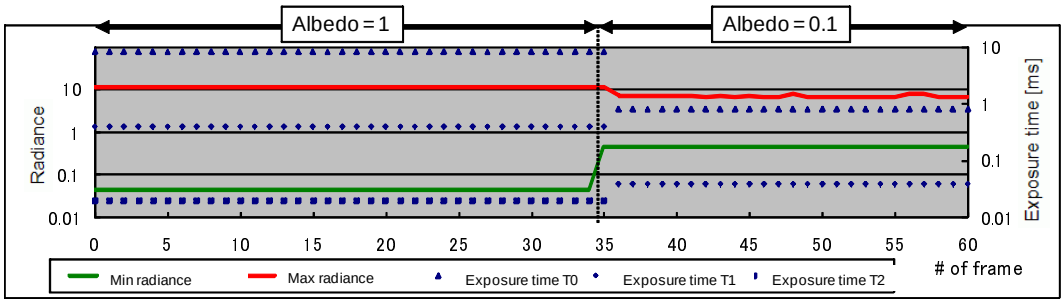
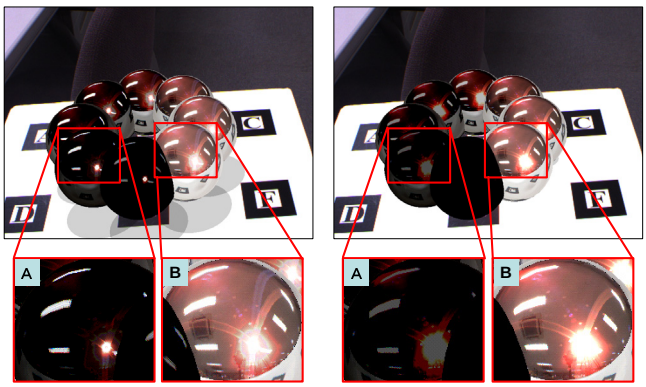


Figure 5: Estimation results of the number of images and exposure times(Effect of change in albedo values).



(a) Image generated using an HDR image (proposed method)

(b) Image generated without an HDR image (conventional method)

Figure 6: Synthesized images with different reflectances of virtual objects.

the camera for lighting environment estimation for different albedo values. The red and green lines in the figure indicate the maximum radiance E_{envmax} and minimum radiance $max(E_{envmin}, E_{armin})$ of the required dynamic range, and the triangle, diamond, and box symbols indicate exposure times. From the Figure 5, we confirm that the exposure times and the number of images are appropriately updated according to the change in virtual scene. For an albedo of 1, the capture time is 96 ms, whereas for an albedo of 0.1, the capture time is 63 ms. In the conventional method, when ten images are captured with exposure times $\frac{256}{2^n}(0 \leq n < 9)$, assuming a static experimental environment, it takes 670 ms to capture the image.

3.3 Synthesized image generation

We synthesized an image that can represent shadow, shading, and reflectance of a virtual object using HDR images. Figure 6 (a) and (b) show synthesized images obtained by rendering a virtual object with and without an HDR image generated by the proposed method, respectively. In the composed images, the virtual objects with albedo values set from 0.01 to 1.0 are rendered. Part A in the figure illustrates the reflectance of a virtual object with a low surface albedo. On the other hand, Part B shows the reflectance of a virtual object with a high surface albedo. We confirm that an inappropriate white part exists on the surface of the virtual object with a low albedo is low in Part A of Figure 6 (b), and the reflectance of the virtual object is represented correctly in the proposed method, as shown in Figure 6 (a). The frame rate is approximately 15 fps for augmented image generation and 3 fps for lighting environment estimation.

4 Conclusion

This paper has proposed a new framework of photorealistic AR that can represent an HDR reflectance by estimating the lighting environment with adaptive HDR image generation. To avoid appearance of inappropriate white and black parts on the surface of a virtual object due to insufficient dynamic range of the camera measuring the lighting environment, the proposed method can generate a synthesized image by HDR rendering and tone mapping. The calculation time can be adjusted by varying the number of images and exposure times for HDR image generation according to the AR environment.

In future studies, to increase the realism of an AR scene, we should construct an AR system using a camera that can capture an HDR image for not only lighting environment estimation but also from user's view.

Acknowledgments This work is supported in part by Core Research for Evolutional Science and Technology(CREST) Program of Japan Science and Technology Agency (JST).

References

[1] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. *Proc. of International Conference on Computer Vision*, Vol. 2:10–17, 2001.

[2] K. Agusanto, L. Li, Z. Chuangui, and N. W. Sing. Photorealistic rendering for augmented reality using environment illumination. *Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 208–216, 2003.

[3] M. Ashikhmin and J. Goyal. A reality check for tone-mapping operators. *ACM Transactions on Applied Perception*, Vol. 3, No. 4:399–411, 2006.

[4] W. Cho and K. Hong. Extending dynamic range of two color images under different exposures. *Proc. of International Conference on Pattern Recognition*, Vol. 4:853–856, 2004.

[5] J. Cohen, C. Tchou, T. Hawkins, and P. E. Debevec. Real-time high-dynamic range texture mapping. *Proc. of Eurographics Rendering Workshop '01*, pages 313–320, 2001.

[6] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. of SIGGRAPH '98*, pages 189–198, 1998.

[7] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Proc. of SIGGRAPH '97*, pages 369–378, 1997.

[8] M. Kanbara and N. Yokoya. Geometric and photometric registration for real-time augmented reality. *Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 278–279, 2004.

[9] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *Proc. of SIGGRAPH '03*, pages 319–325, 2003.

[10] M. Niskanen. View dependent enhancement of the dynamic range of video. *Proc. of International Conference on Pattern Recognition*, Vol. 1:984–987, 2006.

[11] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. E. Debevec. Direct HDR capture of the sun and sky. *Proc. of AFRIGRAPH '04*, pages 145–149, 2004.