NAIST-IS-DD1161016

**Doctoral Dissertation**

# Imaging and Rendering Framework for Photorealistic Mixed-Reality World Exploration

Fumio Okura

March 24, 2014

Department of Information Science
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Fumio Okura

Thesis Committee:

| | |
|---|---|
| Professor Naokazu Yokoya | (Supervisor) |
| Professor Hirokazu Kato | (Co-supervisor) |
| Associate Professor Masayuki Kanbara | (Co-supervisor) |
| Associate Professor Tomokazu Sato | (Co-supervisor) |

# Imaging and Rendering Framework for Photorealistic Mixed-Reality World Exploration*

Fumio Okura

**Abstract**

The exploration of the real world in a virtual simulation has become one of the most important applications of augmented virtuality (AV) in accordance with the popularization of virtual globe applications (e.g., Google Earth), where AV is one aspect of mixed reality (MR) research fields according to the reality-virtuality continuum. A large-scale physical environment can be virtualized using modern three-dimensional (3D) reconstruction techniques, such as vision-based and multi-sensor integration approaches. Image-based rendering provides promising improvements to the appearance of automatically reconstructed 3D models, which usually include errors in 3D shape.

In certain applications that allow an interactive virtual exploration of the real world, virtual objects can be superimposed for the visualization of disaster areas or non-existent buildings. In such applications, the virtual objects have to be rendered frame by frame. Real-time rendering approaches have been studied for photometric registration between real and virtual objects in augmented reality (AR), and have become one of the most important research fields in the area of MR. The goal of these researches is to achieve the same quality as offline rendering, but this has not been achieved yet.

This study proposes a framework for acquiring real-world scenes, and rendering both the real and virtual worlds for photorealistic MR-world exploration.

*Doctoral Dissertation, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1161016, March 24, 2014.

Spherical and/or high-dynamic range (HDR) imaging techniques have been studied for photorealistic virtual-object superimposition using image-based lighting. Most studies have dealt with imaging on the ground, whereas some difficulties remain in spherical aerial imaging for realizing a bird's-eye view, which is used for most virtual globe applications. This thesis first describes two approaches for acquiring full spherical aerial images using an unmanned airship and omnidirectional cameras, where the images can be used for the photorealistic superimposition of virtual objects. In addition to MR-world exploration, the acquired images/videos can be employed for aerial-immersive panoramas, i.e., an extension of Google Street View type applications for a bird's-eye view. To achieve the photorealistic superimposition of virtual objects onto a virtualized real-world, we propose a rendering approach that combines the offline rendering of virtual objects with image-based rendering. The proposed rendering approach takes advantage of the higher quality of offline rendering without the high computational cost of online computer graphics (CG) rendering, i.e., it incurs only the cost of online computations for light-weight image-based rendering.

Using the proposed imaging and rendering frameworks either fully or partly, various MR applications can be developed. We also demonstrate some examples of such MR applications based on the proposed imaging and rendering frameworks.

**Keywords:**

mixed reality, high-dynamic range imaging, omnidirectional imaging, photorealistic rendering, free-viewpoint image generation

# 写実的な複合現実空間生成のための画像取得および レンダリングフレームワーク *

大倉 史生

## 内容梗概

Google Earth 等の Virtual Globe アプリケーションの普及とともに、インタラクティブに視点を移動しながら現実空間を閲覧することは、複合現実感（Mixed Reality: MR）の一分野である Augmented Virtuality (AV) における最も重要な応用の一つとなっている。近年、これらの応用として、事前撮影された画像群を用いた多視点ステレオ等による三次元復元により実空間を仮想化し、効率的に広範囲の情景を提示することが可能となってきた。さらに、自動復元された三次元形状の誤差に起因する見えへの悪影響を軽減するため、イメージベースの画像生成アプローチを取り入れた自由視点画像生成についての研究が盛んに行われている。

インタラクティブに実空間の閲覧を可能とするアプリケーションにおいて、災害や建造物等の景観シミュレーションのために、仮想物体を合成する応用が考えられる。そのためには、提示画像のフレームごとに、リアルタイムのレンダリングにより仮想物体の見えを計算する必要がある。特に拡張現実感（Augmented Reality: AR）において、CG 分野で盛んに研究されてきたリアルタイムレンダリング手法を AR に適用することにより光学的整合性を向上することが試みられてきたが、現在も、これらの手法はオフラインで時間をかけたレンダリング品質には追いついていない。

本論文は、実空間と仮想物体が写実的に融合された複合現実空間を構築するための、画像取得および実・仮想環境のレンダリングのためのフレームワークについて述べる。従来、仮想物体合成のための実空間の撮影手法として、全方位カメラ等を用い、全天球かつ高ダイナミックレンジの画像を取得する手法が実現されてきた。しかし、多くの手法は地上撮影を対象としており、多くの Virtual

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 博士論文, NAIST-IS-DD1161016, 2014 年 3 月 24 日.

Globe アプリケーションで用いられるような空からの視点での全天球撮影は困難であった。そこで、本論文ではまず、複合現実空間生成に適した空撮手法として、全方位カメラと無人飛行船を用いた不可視領域のない全天球映像の生成方式を提案する。撮影された全天球映像は複合現実空間の生成だけでなく、Google Street View に代表される実環境の見回しアプリケーション等を空からの視点に拡張することができる。本論文ではさらに、全天球撮影に基づき仮想化された現実空間に仮想物体を写実的に合成するための、オフラインレンダリングと自由視点画像生成を融合した仮想物体合成アプローチを提案する。本研究で提案する仮想物体合成は、オンライン処理における仮想物体のレンダリングのための計算量を必要とせず、簡略化された自由視点画像生成の処理コストのみで複合現実空間を提示可能とする。

画像取得およびレンダリングフレームワークを部分的に用いることによって、様々なアプリケーションを構築可能である。本論文では、フレームワークの提案だけでなく、それらを用いた複合現実感アプリケーションの構築例についてもあわせて述べる。

キーワード

複合現実感, 高ダイナミックレンジ画像, 全方位画像, 写実的レンダリング, 自由視点画像生成

# Contents

# List of Figures

# List of Tables

# Glossary

1D one dimensional

2D two dimensional

3D three dimensional

ANOVA analysis of variance

AR augmented reality

AV augmented virtuality

CCD charge coupled device

CG computer graphics

CMOS complementary metal oxide semiconductor

DoF degrees of freedom

FOV field of view

GI global illumination

GLSL OpenGL Shading Language

GPS global positioning system

GPU graphics processing unit

HDR high dynamic range

HMD head mounted display

IBL image-based lighting

IBR image-based rendering

LDR low dynamic range

LI local illumination

MBR model-based rendering

MR mixed reality

MTB median threshold bitmap

MVS multi-view stereo

ND neutral density

OMS omnidirectional mutli-camera system

PC personal computer

RANSAC random sample consensus

SfM structure from motion

SNR signal to noise ratio

SSD sum of squared differences

UAV unmanned aerial vehicle

VDTM view-dependent texture mapping

VR virtual reality

# Chapter 1

# Introduction

For thousands of years, human-beings have *virtualized* the real world, and explored the virtualized real world. The map is a typical example of virtualization. One of the oldest maps remains on a rock at Valcamonica, Italy, which was drawn in approximately B.C. 1500 [Bra97]. From the invention of the map, people have virtualized all over the world, and explored maps for learning about the world. In 1827, Joseph Nicéphore Niépce developed the photographic technology, which is a novel methodology for virtualizing the real world. This technology enabled to explore the world with realistic representations. Information science, mixed reality (MR) in particular, has largely extended the role of the real world virtualization. Telepresence [Min80, Ste92] is a primary approach of information science based on virtualization, which is a technique presenting users highly immersive real-world scenes of remote site; that is, a goal of this technique is to offer people a virtual transporter. As a result of the popularization of virtual globe applications [But06] (e.g., Google Earth), we can currently explore the virtualizations not only for learning the world, but also for immersing into the world. Interactive exploration of the real world in a virtual simulation (e.g., telepresence and virtual globes) is now one of the most important uses of augmented virtuality (AV), which is an aspect of MR research fields according to the reality-virtuality continuum [MTUK94], as shown in Figure 1.1. In such applications, a large-scale physical environment can be virtualized using modern three dimensional (3D) reconstruction techniques, such as vision-based [ASS+09, ASSS10, WACS11] and multi-sensor integration approaches [BMOI08, PNF+08]. To improve the appear-

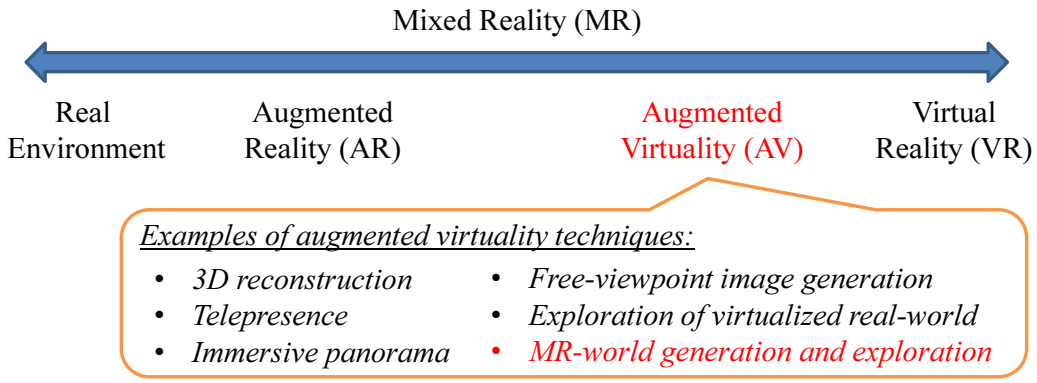



Figure 1.1. Reality-virtuality continuum [MK94] and augmented virtuality techniques.

ance of automatically constructed 3D models, free-viewpoint (sometimes referred to as novel- or arbitrary-viewpoint) image generation using a hybrid of image- and model-based rendering [DTM96, ZC04] provides promising solutions [SSS06].

Human-beings also have *created* countless new worlds. History of creation is older than that of virtualization. A recent study [PHGD+12] figured out that some old rock paintings, in which imaginary or abstract scenes were occasionally depicted, were drawn over 40.8 thousand years ago. In MR fields, the history of creation is also older than that of virtualization. Virtual reality (VR) is an MR technique that creates virtual worlds in the cyberspace, and to present it for users with high immersive values (right edge of reality-virtuality continuum shown in Figure 1.1). The Ultimate Display [Sut65, Sut68], whose concept was proposed in 1965 by Ivan Sutherland, is the roots of VR as well as of the other MR studies. It is the first concept of immersive technology using a head mounted display (HMD), and yet predicted an ultimate goal of VR. In the last few decades, virtual objects are occasionally superimposed onto the real world in real-time. Such technique is referred to as augmented reality (AR) [Azu97, ABB+01], and has recently attracted lots of attention from many creators as well as from industry.

In certain applications that allow an interactive exploration of the virtualized real world, virtual objects can also be superimposed for the visualization of disaster areas or non-existent buildings [Cho09, Gro05, GB08]. In this thesis,

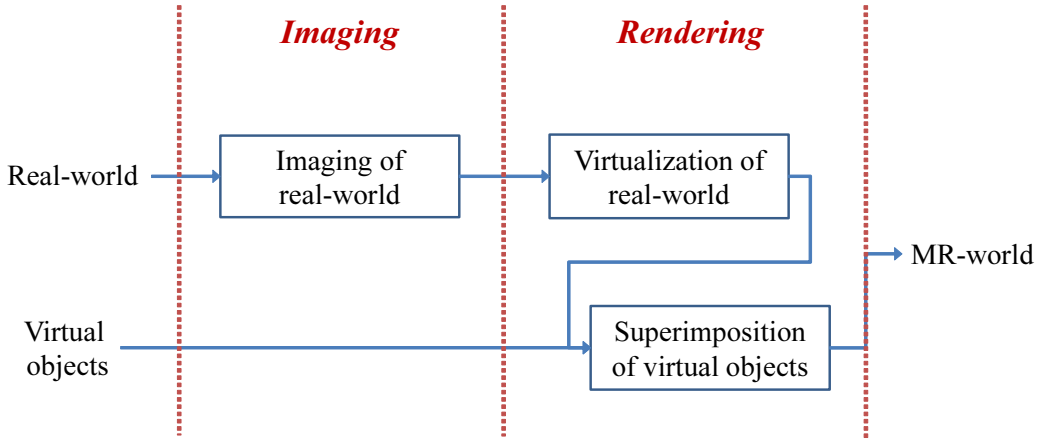



Figure 1.2. Basic pipeline of MR-world generation.

such type of MR applications is referred to as MR-world exploration. Figure 1.2 illustrates the basic pipeline of generating an MR world, which consists of imaging, real-world virtualization, and virtual-object superimposition. We refer to the generation process as MR-world generation.

In MR-world exploration applications, virtual objects are rendered on a frame-by-frame basis, along with virtual-object superimposition in AR applications. A goal of MR applications using virtual-object superimposition is to realize an MR world where virtual objects are superimposed into real scenes as if they actually exist. Photometric registration [KY02] between real and virtual objects is one of the most important factors in realizing photorealistic virtual-object superimposition. It is basically achieved in two steps: estimating the real-world illumination, and rendering virtual objects using the estimated illumination. There are numerous AR/MR studies that have focused on one or both of these steps.

In traditional approaches to real-world illumination estimation, scenes of real environments are acquired directly using mirrored balls [Deb98] or cameras with fisheye lenses [GM00, SSI02, ALCS03]. Some recent studies have been able to successfully estimate the illumination from AR background images without the use of additional cameras or objects [JND12, GRTS12]. However, most studies have dealt with the illumination acquisition from the ground, and there have been some difficulties in acquiring real scenes from the sky that can be used

for realizing fly-through MR-world exploration applications with a photorealistic superimposition of virtual objects. Such fly-through views are used for most virtual globe applications.

This study proposes the use of aerial imaging techniques that overcome the difficulties in realizing fly-through MR-world exploration. For photometric and immersive presentations of real and virtual objects, real-world images are captured through full spherical high dynamic range (HDR) imaging using omnidirectional cameras mounted on an unmanned airship. In addition to MR-world exploration, the acquired images/videos can be employed for aerial immersive panoramas, i.e., an extension of Google Street View type applications for a bird's-eye view.

The process of generating the appearance of virtual objects using an estimated illumination is achieved by employing real-time rendering techniques [HDH03, KY02, GM00, GCHH03, KTM$^+$10, GDS10, KK12, LB12], such as those used in computer graphics (CG) research. The quality of real-time rendering has been improved with advancements in computer technology. However, real-time rendering has not achieved the same quality as offline rendering, which is generally time-consuming. Although real-world illumination conditions and virtual-object reflectance properties can be estimated/determined accurately, part of their information is wasted because real-time rendering cannot represent all physical phenomena. Notably, computational costs become greater when high-quality rendering techniques are used, or when virtual objects consisting of numerous polygons are applied. In these types of situations, it is difficult to perform state-of-the-art real-time rendering techniques on mobile devices. For this reason, *cinema-quality* mobile AR has not been achieved yet. In film-making, on the other hand, CG effects are rendered using high-quality and time-consuming illumination methods, and a large number of manual operations. We contend that the efforts involved in such offline and high-quality processes can be employed to improve virtual-object expression in MR applications.

To achieve the photorealistic superimposition of virtual objects utilizing offline effort, we propose a new rendering approach that combines the offline rendering of virtual objects and free-viewpoint image generation. The proposed rendering approach generates the appearance of virtual objects using free-viewpoint image generation from offline-rendered images with multiple viewpoints to take advan-

tage of the higher quality of offline rendering and without the computational cost of online CG rendering; i.e., it incurs only the cost of the online computations for light-weight free-viewpoint image generation. In addition to aerial spherical images acquired by the proposed imaging techniques, this approach accommodates the input of general real-world scenes with only minor modifications.

The remainder of this thesis is organized as follows. Chapter 2 describes other works related to imaging and rendering techniques for MR-world generation and presentation, followed by descriptions about contribution of this study. Next, the aerial spherical imaging approaches are detailed in Chapter 3, and the proposed rendering approach is introduced in Chapter 4. The proposed imaging and rendering approaches can be employed for various MR applications using parts of the framework with minor modifications. Chapter 5 describes examples of MR applications based on the proposed framework, along with discussions of the methodology and limitations in developing each application. Finally, Chapter 6 summarizes this thesis.

# Chapter 2

# Related Work

## 2.1. Imaging Technique for MR-World Generation

We first discuss the requirements for imaging techniques used to achieve our aim, that is, photorealistic MR-world generation. Since the original invention of photography by Joseph Nicéphore Niépce in 1827, numerous imaging techniques have been developed. Digital monocular cameras are currently widespread and are attached to most mobile devices. However, ordinary monocular imaging is not sufficient for MR-world generation, the requirements of which can be summarized into the following two categories.

- Effective acquisition of large environments
  In most MR-world exploration applications, a real scene is virtualized using images captured from numerous locations and directions. To acquire an entire scene of the target environment, effective imaging techniques are required. Inefficient methods, such as the use of conventional handheld monocular cameras, can easily be used to generate the missing parts of a real-world virtualization drastically lacking user immersion. In addition, such missing parts may cause a lack of illumination of the real world, which negatively affects the appearance of the virtual objects.

- Accurate acquisition of the scene luminance
  The rendering of virtual objects is an estimation process of their surface color. To achieve photorealistic rendering, the surface color should be estimated accurately using the surrounding illumination and reflectance property of the surface. This implies that an accurate acquisition of the real-world illumination should be achieved for photorealistic MR. Unfortunately, ordinary cameras are unable to acquire accurate luminance information of a light source owing to a saturation of the pixel intensity.

According to the above requirements for photorealistic MR-world generation, we will review the existing imaging techniques in terms of the field of view, dynamic range, and capturing location.

### 2.1.1 Field of View

Large field-of-view (FOV) imaging techniques facilitate the effective acquisition of a real scene.

#### Fisheye camera

Traditional imaging systems basically employ a CCD/CMOS array and a lens, the projection model of which can be approximated by a planer perspective projection similar to a pinhole camera. To realize large FOV imaging with such traditional cameras, a fisheye lens, which has a very short focal length, is employed [Miy64]. Fisheye lenses can be employed without any special equipment beyond an ordinary camera and the lens itself, and acquire about 180-degree images[1], as shown in Figure 2.1. Xiong and Turkowski [XT97] proposed a registration method for multiple fisheye cameras to acquire a FOV of larger than 180 degrees. However, in principle, they cannot acquire 360-degree, omnidirectional scenes using a single shot.

[1]Fisheye lenses wider than 180-degree can be developed in principle. In a particular case, Nikon has released a 220-degree fisheye lens (e.g., Fisheye Nikkor 6mm f/5.6 released in 1969).


Figure 2.1. Fisheye camera and acquired image in [XT97].

### Omnidirectional camera using mirrors

Omnidirectional imaging, in which the entire direction of a scene is acquired, is a popular and ultimately large FOV imaging technique [Yag99]. Omnidirectional imaging techniques employing a traditional camera and mirrored surfaces have traditionally been studied. Image-based lighting (IBL), a CG rendering method based on real-world illumination, originally employed two mirrored spheres and two times of capturing [DTM96]. In the field of robotics, particularly for the navigation of mobile robots, imaging systems using a mirror and a video camera have been developed to capture real-time omnidirectional videos [Nay88, YK90, YYY93, Nal96, Nay97]. This type of omnidirectional camera is referred to as catadioptric camera. Such studies have developed mirrored surfaces and realized omnidirectional projection models with a common principal point. However, it is difficult to acquire a large vertical FOV using catadioptric cameras.

### Omnidirectional multi-camera system

In accordance with the downsizing of traditional cameras, omnidirectional multi-camera systems (OMSs) have been developed. OMSs employ multiple cameras synchronized to acquire the entire direction, and combine multiple images into a single omnidirectional image. With a well-considered configuration, an OMS can capture a nearly full sphere with a single shot. In this thesis, such an omnidirectional image covering nearly an entire sphere is referred to as a spherical image. Although most OMSs do not achieve common principal points among multiple cameras, they easily achieve a higher resolution and wider vertical FOV than catadioptric cameras. OMSs have become popular as imaging systems used for Google Street View and certain commercial products used in research, such as Ladybug by Point Grey Research, Inc. (http://www.ptgrey.com/). The growth of OMSs has brought omnidirectional/spherical imaging techniques to the consumer market in recent years. Notably, Ricoh Theta, released by Ricoh (http://theta360.com/) in 2013, is expected to standardize the role of omnidirectional imaging for the general public. This product is equipped with two CMOS sensors with an over 180-degree fisheye lens, and combines two images into a single spherical image. Omnidirectional imaging techniques are widespread, and are expected to become even more popular and important.

### The best large FOV imaging method for MR-world generation

Table 2.1 summarizes major approaches for large FOV imaging. With the recent growth of OMS, high-resolution spherical imaging using OMS has become not so costly compared to former approaches. Because most OMSs do not achieve common principal points among multiple cameras, they cause misalignments on panoramic spherical images. Although a dynamic panoramic stitching method can be applied similar to a real-world exploration application developed by Uyttendaele et al. [UCK+04] to reduce the visual artifacts due to the misalignment, it is difficult to compensate the difference of the principal points.

In terms of MR-world generation, which is expected to be employed in large environments, the misalignment problem can be negligible small. The distance between the principal points of neighboring cameras in Ladybug2, which is a fa-

mous OMS developed by Point Grey Research, is approximately 2cm. In the case the distance to the real scene is 15m, misalignment of the scene on the spherical image is 0.07°, which is less than one pixel of the maximum resolution of the camera. It intends that the misalignment is ignorable if the target environment consists of long-distant view, which is a general situation in outdoor shooting. Although it required to pay careful attention to the misalignment issues, imaging using OMS is a promising approach to fulfill a requirement of imaging techniques for MR-world generation; that is, effective scene acquisition of large environments.

### 2.1.2 Dynamic Range

To employ real-world images as an illumination environment for rendering virtual objects, the images should be unsaturated [DM97]. Unfortunately, the dynamic range of many outdoor scenes during the daytime is too high for a proper capture using standard 8-bit (low dynamic range, or LDR) cameras, as sunlight can reach approximately $2^{17}$-times brighter levels than darker areas of the sky or cloud cover [STJ+04]. HDR imaging techniques, which acquire higher than 8-bit real-world images, have been studied to overcome this problem.

Table 2.1. Large FOV imaging approaches.

| | Fisheye camera | Omnidirectional camera w/ mirrors | Omnidirectional multi-camera system |
|---|---|---|---|
| One-shot 360° imaging | × | √ (narrow vertical FOV) | √ (near full sphere) |
| Resolution | × | × | √ |
| Cost | √ | √ | × (formerly) √ (recently) |
| Common principal point | √ | √ | × |
| Weight | $< 1$ kg | $< 1$ kg | $> 1$ kg |

### HDR camera

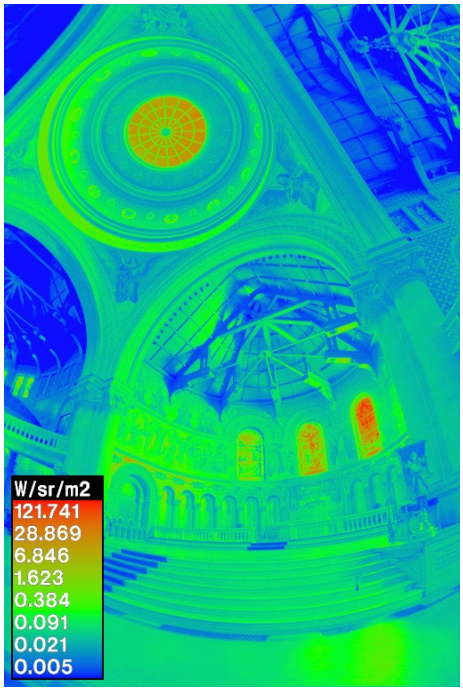
Some HDR cameras equipped with special CCD/CMOS sensors with a high-quality A/D converter are available for film-making, and achieve 10- or 12-bit/channel video recording. ViewPLUS, Inc. (http://www.viewplus.co.jp/) has released an 18-bit HDR video camera, called Xviii, for research use. Spherical HDR cameras have also been developed for use in virtual reality. SpheronVR (http://www.spheron.com/) has released a 13,000 pixel $\times$ 5,300 pixel full spherical panorama camera, the bit-depth of which reaches a dynamic range of $10^8 : 1$; however, the acquisition of a single shot takes several minutes owing to the use of a line-scan sensor.

A spherical HDR camera that acquires a spherical image in a single shot, can be a suitable imaging technique for MR-world generation, in terms of its effectiveness and accuracy. However, it is so far difficult to find such a special HDR camera.

### Multi-exposure imaging

An HDR imaging method using LDR images taken by multiple exposures (multi-exposure images), shown in Figure 2.2, is widely used to capture scenes with a high dynamic range using commodity camera hardware. Exposure of the digital camera, the amount of the light incoming to a sensor, is determined by changing exposure time, aperture size, and sensor gain. Multi-exposure images are captured by changing exposure so as to capture both bright and dark portions. An HDR image is generated by appropriately combine these images. This approach was originally proposed by Debevec et al. in 1997 [DM97], and multi-exposure imaging is currently being implemented in most commercial still-cameras as an auto-bracketing function. It should be noted that the response curve (sometimes referred to as the response function) of such cameras is basically non-linear, that is, the pixel intensities do not vary linearly in accordance with the physical luminance. To successfully generate an HDR image from multi-exposure images, the response curve should be calibrated and linearized. Two representative calibration techniques were developed by Debevec et al. [DM97] and Mitsunaga et al. [MN99]. Debevec et al. [DM97] approximated the camera response using a logarithmic function, whereas the other method [MN99] employs a polynomial

(a) Multi-exposure images.




(b) Recovered HDR radiance.

Figure 2.2. HDR radiance reconstruction from multi-exposure images [DM97].

function. The former method [DM97] is currently being implemented as a software library for HDR image composition [DT01]. For multi-exposure imaging, exposure time should be changed rather than other parameters to avoid change of depth of field and increase of noises. Because changeable exposure time is often restricted by the hardware, the aperture size is also changed in some practical examples of multi-exposure imaging [DM97].

When images are captured during motion, however, a misalignment will naturally occur. The alignment of multi-exposure images is one of the most important issues in HDR imaging research. A method proposed by Kang et al. [KUWS03] tracks the scene movement using the Lucas and Kanade tracker [ST94], and then registers the images using a warping technique. To track the movement, a more sophisticated method employs a median threshold bitmap (MTB) [War03], which is robust under large exposure changes. This MTB-based method has been improved to realize a more robust alignment [Gro06, JLW08]. However, large misalignments remain difficult to fix using alignment methods. For this reason, it is better to use fewer multi-exposure images to generate HDR images, and to ensure that the camera velocity and frame rate remain constant.

#### The best HDR imaging method for MR-world generation

As discussed in Section 2.1.1, using large FOV cameras, OMSs in particular, is a suitable solution for effective scene acquisition of a large environment. It is currently difficult to find special OMSs equipped with HDR sensors. Therefore, multi-exposure imaging is a practical solution for HDR spherical imaging. Note that, the misalignment between the multi-exposure images should be taken into account when the camera is not fixed during multi-exposure imaging, which is expected to be a common situation in MR-world generation.

### 2.1.3 Capturing Location

The large FOV and/or HDR imaging techniques described above are employed under various situations. In this section, we categorize these imaging systems in terms of the capturing locations: from the ground and the sky.

### Imaging from the ground

Most imaging techniques are intended for use on the ground. Omnidirectional imaging is often employed in immersive panorama-based real-world exploration applications as detailed in Section 2.2.1. Certain immersive panorama applications acquire spherical images/videos from a vehicle-mounted OMS [ADF+10]. In addition to a vehicle-mounted capture, spherical images can be captured through various settings using a tripod- or human-mounted OMS [UCK+04].

Unless an OMS realizing almost full spherical imaging is employed, missing areas often remain in a single-capture image, as shown in Figure 2.3. Such missing areas are caused by limitations in the FOV of the OMS and/or occlusions of the FOV by persons or vehicles mounting the camera. Kawai et al. have dealt with the completion of such occlusions, particularly for spherical videos captured on the ground [KMSY10]. In the method in [KMSY10], missing areas in the ground portion of spherical video sequences were filled in by estimating the intensities of the missing areas based on the intensities of the same locations taken from other viewpoints. The result of the method in [KMSY10] is shown in Figure 2.3.

When developing immersive panorama applications in crowded areas (e.g., an underground city), a large number of pedestrians occlude the background scenery. In the application for an underground city visualization developed by Arai et al. [AHK+10] (see Figure 2.4), time-series images captured at fixed locations were unified such that only the intensity of the background scenery is used based on the temporal median intensity. Because some pedestrians may remain in a temporal-median based approach, an image inpainting method [KSY09] was also employed to completely remove them.

HDR spherical images can be captured through multi-exposure imaging from the ground. Such images have been traditionally captured for IBL using mirrored spheres [DTM96] mounted on a tripod. Naturally, we can employ an OMS for HDR spherical image capturing in a straightforward manner. Appendix A describes a simple approach used to acquire a spherical HDR image from the ground. Figure 2.5 shows an example of an HDR spherical image captured in an underground city, whose radiance is transformed into an 8-bit image using a tone-mapping method [RSSF02]. In addition to IBL, these images help improve the visibility of the scenery in immersive panorama applications.

(a) Spherical image with a missing area.

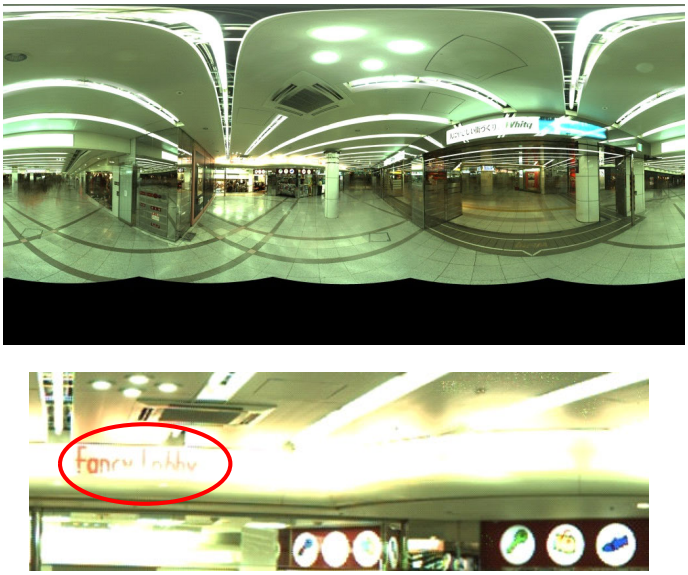
(b) Complete spherical image.

Figure 2.3. Spherical image completion for missing areas in the ground portion [KMSY10].

(a) Spherical image with pedestrians.

(b) Spherical image with pedestrians removed.

Figure 2.4. Pedestrian removal in spherical images [AHK+10].



(a) A multi-exposure image.

(b) HDR spherical image whose radiance is compressed through the method in [RSSF02]

Figure 2.5. An image from a multi-exposure spherical image combined with a generated HDR spherical image.

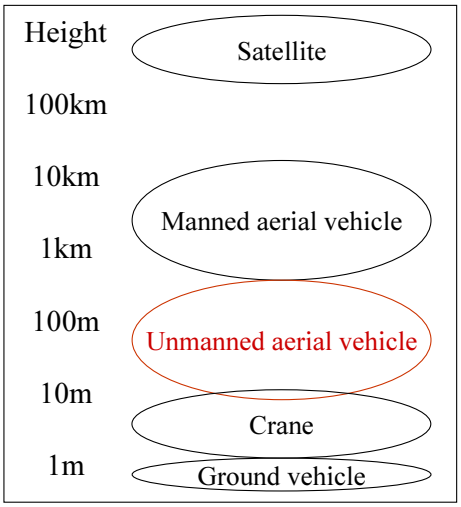



Figure 2.6. Aircraft altitude.

### Aerial imaging

Aerial imaging using monocular cameras is traditionally used for various applications (e.g., satellite maps and 3D reconstruction). Most map applications employ a bird's-eye viewpoint using aerial images, rather than a viewpoint from the ground, because a bird's-eye view is advantageous in grasping large-area scenes at a glance. Aerial images are usually captured from a satellite or airplane at an altitude of over 1 km, as illustrated in Figure 2.6. Unmanned aerial vehicles (UAV) have recently attracted attention for low-altitude aerial imaging [HJD+04, GMG+08], which can acquire higher resolution images than high-altitude aerial imaging. In particular, lighter-than-air UAVs (e.g., unmanned airships), which do not require fuel to float, are often used for low-altitude aerial imaging [PARJ+06, FYS+08, HJSL04]. The benefits of lighter-than-air UAVs, i.e., fuel-efficiency and portability, are shown in Table 2.2.

Although aerial-imaging techniques have been widely used, there are few studies dealing with omnidirectional/spherical imaging from the sky. Omnidirectional cameras are occasionally mounted on UAVs for visual servo (i.e., vehicle control) tasks [HS03, MCMO10]. In addition to vehicle control, omnidirectional aerial imaging can be used for immersive panoramas, effective 3D reconstruction, and

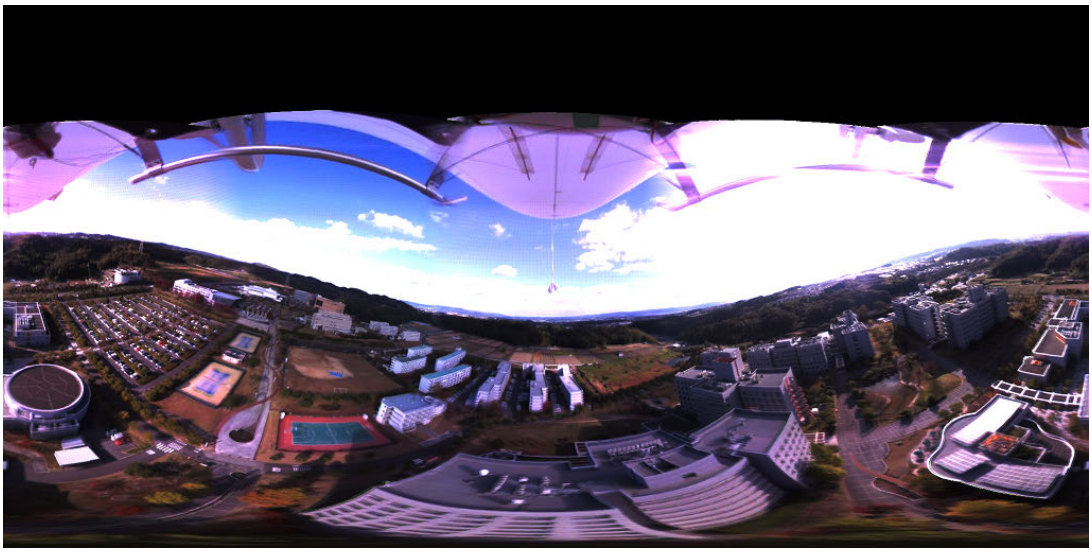


Figure 2.7. Spherical aerial image captured from an OMS mounted on bottom of airship.

virtual object superimposition. Figure 2.7 shows a spherical image captured from the sky by an OMS mounted beneath an unmanned airship. A problem tends to occur in spherical aerial images acquired using an OMS, i.e., the occurrence of missing areas. While this problem is also typical in images taken from the ground, it is a difficulty inherent to aerial imaging. Such missing areas are caused by limitations in the FOV of an omnidirectional camera and/or occlusions of the FOV by the hardware used. The primary occlusion shown in Figure 2.7 is from the aircraft on which the camera is mounted. Such areas subtract from the immer-

Table 2.2. Comparison of different aircrafts.

| | Airplane | Helicopter | Unmanned airship |
|---|---|---|---|
| Altitude | High (over 1 km) | Middle | Low (lower than 150 m) |
| Speed | High | Middle | Low |
| Fuel-efficiency | × | × | √ (lighter than air) |
| Weight | > 1000 kg | > 100 kg | Approx. 10 kg (used in our study) |

sive value of the image and disqualify it from use in rendering system using IBL. While some studies have dealt with missing areas in spherical images [KMSY10], the solutions achieved are not intended for use in aerial imaging.

**Imaging issues for MR-world generation**

There are large variations of the target environment of MR-world generation; it includes both ground and aerial scenarios. Imaging from the ground is employed for street-view applications; while most virtual globe applications employ aerial imaging. Table 2.3 simply summarizes the problem in aerial imaging in terms of the capturing location. Although the intensity of the sky is important for immersion and virtual object rendering in MR-world exploration applications, spherical aerial imaging using OMSs causes missing areas in the sky portion in the spherical images. This fact indicates that we should overcome the problems in aerial imaging to the effective and accurate MR-world generation.

## 2.2. Rendering Technique for Real and Virtual Objects

Numerous rendering approaches have been proposed. This section classifies the rendering approaches for both real and virtual objects according to the flow of the photorealistic MR-world generation, illustrated in Figure 1.2. In a rendering pipeline used in photorealistic MR-world generation, real-world scenes, acquired from the imaging technique described in Section 2.1, are first utilized for virtualization. Virtual objects are then superimposed into the real-world virtualization using CG-rendering techniques.

Table 2.3. Capturing location of large FOV imaging.

| | Spherical imaging w/ missing area | Full spherical imaging |
|---|---|---|
| Ground | $\surd$ | $\surd$ (completion of ground portion [KMSY10]) |
| Aerial | $\surd$ | $\times$ |

### 2.2.1 Real-World Virtualization

Real-world scenes acquired at numerous locations by numerous imaging techniques have been employed for real-world virtualization. The aim of the virtualization process is to develop a tool for exploring the real world. Web-map applications including Google Street View and Google Earth, which employ virtualization techniques, have recently become popular.

Virtualization approaches can be divided into the following two categories in terms of a changeable viewpoint:

1. Immersive panoramas
   An immersive panorama is a real-world rendered image allowing the user to change their viewing direction rather than their viewpoint. Immersive panoramas are employed in Google Street View.

2. Free-viewpoint image generation
   Using free-viewpoint image generation, the user can freely configure both their viewpoint and direction. Virtual globe applications (e.g., Google Earth) employ this technique.

#### Immersive panoramas

An immersive panorama is an image representation using spherical images, which are acquired as described in Section 2.1. It enables users to change their viewing direction from a fixed location where the image was captured. Immersive panoramas are currently very popular with Google Street View [Vin07, ADF+10], as shown in Figure 2.8.

The primary system used for creating an immersive panorama is QuickTime VR [Che95], which employs spherical panoramic images composed from monocular camera images captured from various directions at a single location. A panoramic image is converted into a planar-perspective image, as shown in Figure 2.9, whose viewing direction can be configured freely using the Apple QuickTime framework.

Immersive panorama applications can become more *immersive* using an HMD or large screen, as shown in Figure 2.10. Onoe et al. [OYTY98] proposed a real-time perspective-conversion system using spherical images captured from omnidi-


Figure 2.8. Google Street View [Vin07, ADF+10]

Figure 2.9. QuickTime VR [Che95]

(a) Head-mounted display.

(b) Immersive screen with a motion platform [HKY09].

Figure 2.10. Immersive displays.

rectional cameras, and developed an immersive system using an HMD. An HMD, which is equipped with an orientation sensor, displays a view-direction-dependent perspective image converted from a spherical image. Unlike QuickTime VR, this system can present video sequences acquired during motion by an omnidirectional camera. A system developed by Ikeda et al. [ISKY04] employs a locomotion interface and large screens to realize the sense of walking along a captured path of a spherical video sequence. The employment of motion platforms [HKY09], as shown in Figure 2.10(b), can also contribute to increase the immersive value for spherical videos captured in a shaky environment, such as on a roller coaster.

### Free-viewpoint image generation

Free-viewpoint image generation (sometimes referred to as arbitrary- or novel-viewpoint image generation) is a technique for generating freely configurable views from multiple images captured at multiple locations, and was developed in the fields of CG and computer vision. Free-viewpoint image generation approaches can be categorized according to the geometry-image continuum [KSA00] (see Figure 2.11), which consists of model-based rendering (MBR), image-based rendering (IBR), and a hybrid of the two.

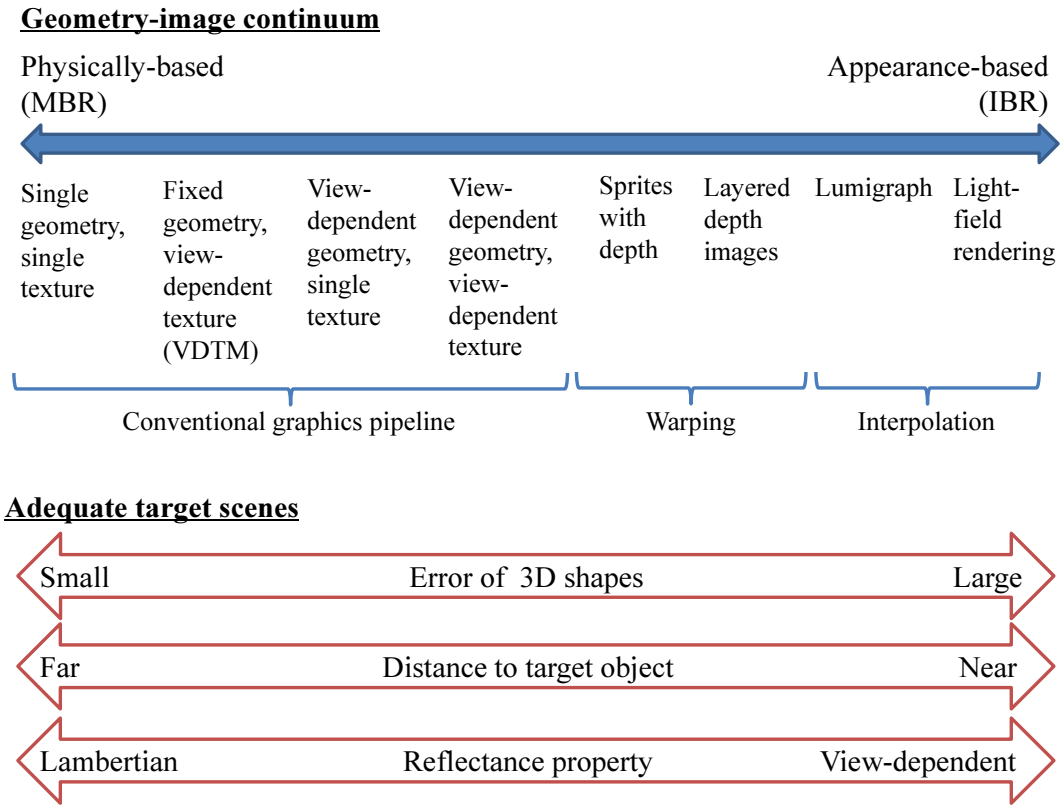



Figure 2.11. Geometry-image continuum [KSA00] and adequate scenes for free-viewpoint image generation methods.

An MBR approach is a traditional computer graphics/vision pipeline that first reconstructs 3D shapes of real environments, and then maps images of the real environment onto the 3D shapes as textures. Some web-maps, including Google Earth, present textured 3D models of buildings. As the most basic MBR approach, handcrafted 3D models can be utilized; for example, a system developed by Shimamura et al. [STYY00] realizes small changes in viewpoint by mapping an omnidirectional image onto a hand-made 3D model. With the recent growth of computer vision techniques, the 3D shape of the real world can be automatically reconstructed by using multiple images and/or laser scanners. 3D shapes can be acquired from the environments ranging from a small desktop space [IKH+11] to a large outdoor environment [MAW+07]. Laser scanners are

occasionally employed for the automatic 3D modeling of a large outdoor environment [BMOI08, NCS+09]. In addition, there are numerous approaches using vision-based 3D reconstruction methods that employ multiple images captured at multiple locations. In vision-based approaches, the position and orientation of a camera capturing multiple images are first estimated through structure-from-motion with bundle adjustment techniques [TMHF99]. Bundle adjustment has recently been speeded up dramatically, and can accommodate large outdoor environments [ASS+09, LA09, ASSS10, WACS11]. A multi-view stereo method is then used to reconstruct a 3D model with a dense surface from multiple images based on their estimated position and orientation [PNF+08, SY10, JP11]. These large-scale reconstruction approaches have been employed for recent web-maps. The quality of the free-viewpoint images generated through MBR is directly affected by the accuracy of the 3D shapes, that is, unnatural distortions or missing areas in such views are easily exposed.

On the other hand, IBR generates free-viewpoint images without the use of explicit 3D shapes. The ultimate form of IBR is to acquire and represent all possible light rays from all positions and directions, where the rays are described as having a plenoptic function [MB95]. To interpolate light rays from a limited number of image acquisitions, view interpolation techniques have been studied, such as view morphing [MD96], light-field rendering [LH96, NTKH97, DLD12], and lumigraph rendering [GGSC96, BBM+01]. Light-field-based representations have also been recently employed for 3D scene reconstruction for a complex environment [KZP+13]. IBR is currently being employed in numerous applications [SCK07]. A multi-perspective panorama [ZFPW03, KCSC10], which is patched together from parts of multiple images, was developed to interpolate the viewpoint from immersive panorama applications, such as Microsoft Street Slide [KCSC10]. Hori et al. [HKY10] proposed a free-viewpoint stereo image generation system utilizing a multi-perspective panorama approach. Although IBR reduces missing areas in the resulting images, this approach requires images captured at a large number of locations and directions. A recent study reported that the amount of perceived distortion increases along with the distance between the user's viewpoint and position where the image was captured [VCL+11, VRC+13].

In recent years, the main approach for free-viewpoint image generation has

been the use of hybrid techniques that combine MBR and IBR with the goal of resolving their respective problems. The primary approach of hybrid free-viewpoint generation is view-dependent texture mapping (VDTM), which was originally proposed by Debevec et al. [DTM96]. This technique selects and blends multiple textures acquired from multiple cameras, and then maps the blended textures onto 3D models. There have been various improvements in VDTM, with focus on the quality of image generation, the calculation cost, and/or specific applications [DYB98, TKIS00, AS02, PDG05]. Although the textures acquired at the proper positions are selected in VDTM approaches, large errors in the 3D shapes still cause large distortions in the resultant free-viewpoint images. State-of-the-art hybrid rendering methods appropriately transform 3D shapes depending on the position of the virtual viewpoint, while also selecting and blending the textures [IHA02,SY10]. Note that the 3D shapes are often transformed based on the appearance of object shape in two dimensional (2D) images (e.g., object silhouette [CSD11] and region segments [CDSHD13]) rather than photo consistency, which is usually used in 3D reconstruction for MBR. These new hybrid-rendering approaches are referred to as view-dependent geometry and texture, and have been employed in some recent applications [OUSY13]. Although 3D shapes may include large errors, these approaches improve the quality of free-viewpoint images, (e.g., when there are missing regions in the 3D shapes).

Here, we describe a guideline for selecting a free-viewpoint generation method in terms of the target environment. Suitable approaches should be selected depending on the errors of 3D shapes, the distance to the scenes, and the reflectance property of the target environment as shown in the bottom of Figure 2.11. In the case that the target environment can be approximated as Lambertian, and yet the error of the 3D shape is small, physically-based rendering approaches, such as VDTM, generate reasonable results. On the other hands, if the 3D shape is poorly reconstructed or the distance to target objects from the viewpoint is small, appearance-based approaches should generate a result preferable to physically-based one. For particularly complex environments such as scenes from near the tree or objects that have complex reflectance, some hybrid-rendering methods utilizing complex processes basically improve the quality of the resultant images at the cost of heavy computation [SKG+12, KLS+13].

#### Real-world virtualization approach for MR-world exploration

We basically deal with MR-world exploration based on both immersive panorama and free-viewpoint image generation. When the OMSs are employed for the imaging, the different between two approaches is whether a free-viewpoint image generation technique is employed. If not, immersive panorama applications are developed. Free-viewpoint image generation provides viewpoint changes, but degrades the quality of the resultant images; thus, immersive panorama is still valuable for applications that should not display debased scenes.

### 2.2.2 Superimposition of Virtual Objects

Virtual objects are sometimes superimposed on a real-world virtualization [Cho09, GB08, Gro05] in certain applications such as landscape simulations. Google Earth has provided a framework for superimposing CG models using Google SketchUp [Cho09] (acquired by Trimble Navigation Ltd. in 2012), which has been used by numerous visualization researchers [Wol08, XBF+09]. To realize changes in viewpoint as in SketchUp, virtual objects must be rendered frame by frame. It is similar process in real-time AR, which superimposes virtual objects onto a real-scene acquired in real-time. The issues inherent to the virtual-object superimposition of a real-world virtualization, specifically those that provide changes in viewpoint, are also common in AR applications, i.e., both require real-time rendering techniques. In real-time AR, the virtual objects are superimposed through the following steps: 1) illumination estimation of the real world, and 2) the rendering of virtual objects.

#### Illumination estimation of the real world

Illumination estimation from monocular images has been studied in the CG research field [NHIN86, SSI03], and some recent AR studies have achieved real-time estimation [JND12, GRTS12]. On the other hand, MR-world generation through the virtualization of the real world often employs imaging from large FOV cameras, as discussed in Section 2.1.1. For such situations, we can directly utilize the luminance acquired by a camera for the illumination environment. Some systems for virtual object superimposition employ mirrored balls [Deb98, KY02],

fisheye cameras [KY02, SSI02, ALCS03], or omnidirectional cameras to acquire a real-world illumination.

### Rendering of virtual objects

Numerous rendering methods have been proposed in the fields of CG and AR [Cro77, WPF90, HLHS03], and can be divided into two categories: local illumination (LI) and global illumination (GI) rendering. LI rendering does not simulate multiple reflections (interreflection) on the surface of CG objects, whereas GI rendering realizes photorealistic rendering by calculating the interreflection between multiple object surfaces, at the expense of the calculation cost.

[Local illumination (LI) rendering]
Owing to its lightweight computation, LI rendering has been widely used for real-time applications including real-time AR. To realize shadow and shade presentations, a shadow map technique is often employed. For AR application, this technique has been expanded to render multiple virtual objects [GM00], soft shadows [GCHH03], and dynamic objects [KY02]. In contrast, Haller et al. [HDH03] utilize shadow volumes to achieve faster rendering.

[Global illumination (GI) rendering]
The photorealistic appearance of a virtual object can be generated through GI rendering, such as photon mapping [Jen96]. There have been numerous GI rendering methods for offline (i.e., non-real-time) virtual-object superimposition onto real-world scenes, such as in movie special effects. Fournier et al. [FGR93] utilized Progressive Radiosity [CCWG88] for virtual-object superimposition. IBL, an illumination acquisition and rendering framework using a large-FOV imaging technique proposed by Debevec [Deb98], has often been employed in film-making. Although the original IBL employs ray-tracing [Gla89] using an illumination environment acquired by mirrored balls, some recent rendering software using IBL employs other GI techniques. These offline GI-rendering techniques are available as libraries [War94, PH10] and commercial software. In the past, however, it was difficult to employ such techniques for AR owing to their heavy computational requirements.

Some AR studies before early 2000s realize the near real-time GI rendering of virtual objects using an image synthesis from the appearance preliminarily rendered in various illumination environments [SHK+05,NSD94], where the viewpoints of the user and virtual objects must be static. With new advancements in computer technology, recent studies on photometric AR have realized real-time GI rendering in dynamic environments, including the use of a pre-computed radiance transfer [GDS10], differential rendering [KTM+10], real-time ray-tracing [KK12], and reflective shadow maps [LB12]. These approaches employ state-of-the-art real-time rendering methods, which have been studied in the CG field, for AR virtual object representation. However, real-time GI-rendering methods have not achieved the same quality as offline rendering, and their heavy computational requirements are too costly for mobile AR applications. Therefore, cinema-quality mobile AR has yet to be achieved. Even in film-making, for example, CG models are rendered using high-quality illumination, time-consuming methods, and many manual operations.

**Virtual object superimposition for MR-world exploration**

Unlike real-time AR applications, it is expected that MR-world exploration does not required to deal with environment with illumination change, because it preliminarily virtualize the real-scene and the illumination is static as shown in Table 2.4. In addition, the real-world illumination can be directly acquired by OMSs. It is possible that a virtual object superimposition method suitable for MR-world exploration applications can be developed; in which the method realizes highly photorealistic image synthesis, although it does not accommodate the illumination change.

## 2.3. Contribution of This Study

Figure 2.12 illustrates the concrete flow of the proposed framework. As discussed in Section 2.1, novel imaging techniques are required for photorealistic MR-world generation using aerial images. That is, although spherical HDR images can be acquired from the ground, spherical aerial HDR imaging should be developed. Based on the discussions in Section 2.2, we propose a rendering framework that

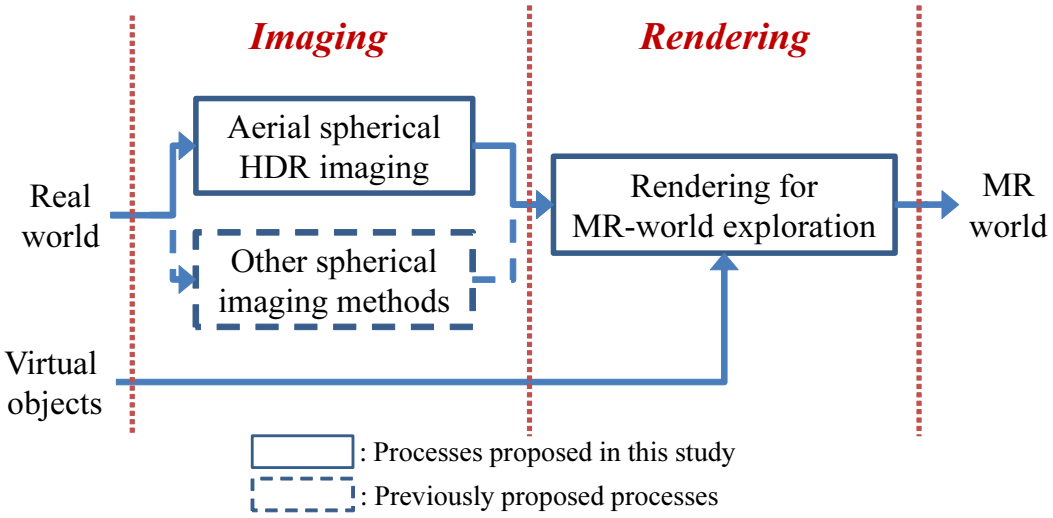



Figure 2.12. The proposed flow in imaging and rendering frameworks.

enables the user to configure their viewpoint freely in a static MR-world, where both real and virtual objects are presented photorealistically. The proposed rendering method can be employed for spherical images captured from both the ground and the sky, as summarized in Table 2.5.

This study proposes two different imaging methods: 1) spherical aerial HDR image generation from a spherical image with missing areas captured from an OMS, and 2) a new imaging technique using two OMSs mounted on the top

Table 2.4. Goals of virtual object superimposition for MR-world exploration. $\surd$: required, $\times$: not required.

| | Real-time (usual) AR | MR-world exploration |
|---|---|---|
| Photorealistic representation | $\surd$ | $\surd$ |
| Illumination estimation of the real world | $\surd$ | $\times$ |
| Accommodating real-time illumination change | $\surd$ | $\times$ |

and bottom of a UAV. The former approach can be employed for images already captured in an ordinary way using an OMS at the expense of the accuracy of the scene luminance. The latter technique requires a special hardware configuration to acquire the scene luminance with high accuracy. It should be noted that the imaging process is done offline. Details of the proposed imaging approaches are described in Chapter 3.

The proposed rendering method is divided into offline and online processes to take advantage of high-quality offline rendering for the real-time representation of the MR-world. In the offline process, real scenes are virtualized using a hybrid free-viewpoint image generation approach, which utilizes automatically reconstructed large-scale 3D models. In addition, the virtual objects are rendered through a novel real-time rendering method that can employ any offline rendering method, that is, scenes that can be rendered photorealistically to the maximum extent. Therefore, strict illumination settings, time-consuming rendering, and any manual adjustment can be included. Offline-rendered images are transformed through real-time free-viewpoint image generation, which preserves the high-quality textures. Details of the proposed rendering framework are described in Chapter 4.

In addition to the MR-world exploration using the proposed imaging and rendering framework, we also show that our framework can be employed in various applications with minor modifications of its implementation, as well as some performance evaluations. Figure 2.13 illustrates the flow of following three applications using part of the proposed framework:

1. HDR immersive panorama (see Section 5.2): Full spherical HDR images can be employed in immersive panorama applications such as Google Street View, by applying perspective projection of the sphere-mapped spherical

Table 2.5. Intended environment of proposed framework.

| | Ground | Sky |
|---|---|---|
| Previous full spherical imaging | $\surd$ | $\times$ |
| Proposed full spherical imaging | $\times$ | $\surd$ |
| Proposed rendering for MR-world exploration | $\surd$ | $\surd$ |

images.

2. Augmented immersive panorama (see Section 5.3): By directly superimposing virtual objects into full spherical aerial images without employing free-viewpoint image generation, the augmented spherical images can be used for augmented immersive panorama applications. This application does not provide a change in viewpoint other than at the location where the images were initially captured.

3. Real-time AR (see Section 5.4): By using real-world scenes captured in real-time, instead of a real-world virtualization, the proposed rendering framework can be employed for (video see-through) real-time AR applications.

4. MR-world interactive exploration (see Section 5.5): When the proposed framework is fully employed, the produced application realizes the capability of interactive MR-world exploration. Section 5.5 introduces two applications based on the implementation described in Section 4.3.

Table 2.6 summarizes the features of MR applications. As shown in the table, our study covers broad application areas. Our applications with application-specific discussions and evaluations, except for MR-world exploration detailed in Chapter 4, are described in Chapter 5.

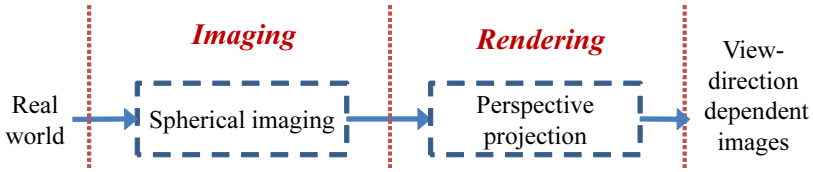
The contributions of this study are summarized as follows:

- Proposing a couple of the world's first full spherical aerial HDR imaging methods with practical examples.

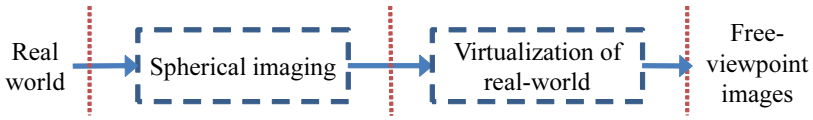
- Proposing a photorealistic and light-weight rendering method for MR-world exploration based on free-viewpoint image generation and offline superimposition of virtual objects.

- Developing MR applications based on the proposed framework, in which the applications can be used for various uses.

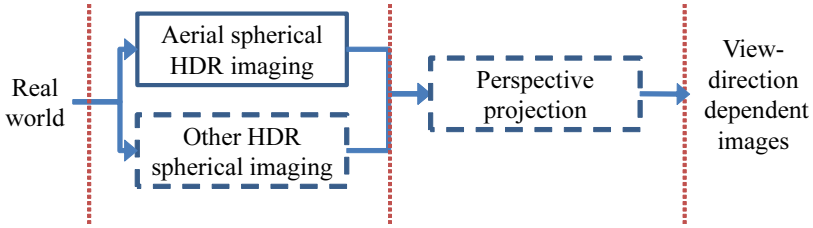
The remainder of this thesis describes the proposed imaging and rendering frameworks, as well as the developed applications.



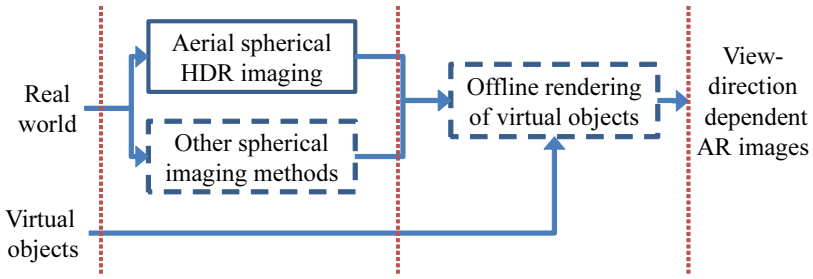

(a) Immersive panorama.



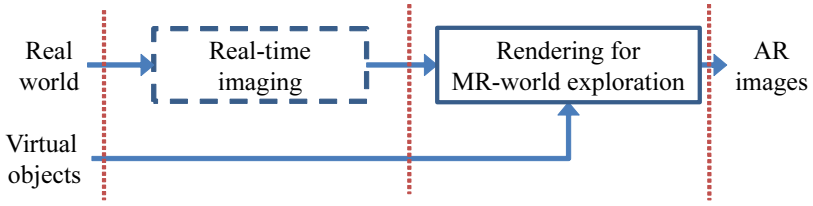

(b) Free-viewpoint image generation.




(c) HDR immersive panorama.




(d) Augmented immersive panorama.




(e) Real-time AR.

Figure 2.13. Applications partially using proposed imaging and rendering frameworks.

Table 2.6. Classification of MR applications. $\surd$: available, †: depending on application, $\times$: not available.

| | | Viewpoint change | HDR | Virtual object | Real scene |
|---|---|---|---|---|---|
| Previous applications | Immersive panorama | $\times$ | $\times$ | $\times$ | Offline |
| | Free-viewpoint image generation | $\surd$ | $\times$ | $\times$ | Offline |
| Applications using proposed framework | HDR immersive panorama | $\times$ | $\surd$ | $\times$ | Offline |
| | Augmented immersive panorama | $\times$ | † | $\surd$ | Offline |
| | Real-time AR | $\surd$ | † | $\surd$ | Real-time |
| | MR-world exploration | $\surd$ | † | $\surd$ | Offline |

# Chapter 3

# Full Spherical Aerial High Dynamic Range Imaging

## 3.1. Overview

In this chapter, we propose two imaging techniques for acquiring full spherical aerial HDR images. To realize photorealistic MR-world generation, large-scale real-world scenes should be effectively and accurately captured. As discussed in Section 2.1, some studies have dealt with spherical HDR imaging from the ground; however, there has been no consideration regarding aerial imaging. The proposed aerial-imaging system resolves a problem concerning missing areas that commonly occur when spherical images are acquired using OMSs.

Missing areas, as shown in Figure 2.7, appear in all single-capture spherical images. Such areas are caused by limitations in the FOV of the OMS and/or occlusions from the vehicles mounting the OMS. In Figure 2.7, the primary occlusion is the aircraft upon which the camera is mounted. Such areas decrease the immersive value of the image and disqualify it from use in photorealistic rendering systems. Although there is a study dealing with missing areas on spherical images acquired from the ground [KMSY10], it is difficult to employ it for aerial imaging because background scenes of missing areas, most of which occludes the sky, cannot be captured by changing the viewpoint. These missing areas can be captured in theory, such as by performing inverted flight; however, these approaches are

not practical in most cases.

This study proposes two approaches overcoming this problem:

1. Spherical aerial image completion using a sky model
   **Input**: A spherical images captured in an ordinary manner using an OMS mounted at the bottom the aerial vehicle.
   **Output**: Full spherical HDR images.
   **Summary**: Missing areas in the spherical images are completed based on the All Sky Model [IKMN04], a statistical model for the luminance and radiance of the sky. Although generated luminance may not reliable, this approach can be used for existing resources captured by ordinary OMS aerial imaging. Note that this approach is utilized by inputting multiple images captured at multiple locations to estimate the captured orientation using vision-based camera pose estimation methods.

2. Full spherical aerial HDR imaging using two OMSs
   **Input**: Multi-exposure spherical images captured using two OMSs from the top and the bottom of an aerial vehicle.
   **Output**: Full spherical HDR images.
   **Summary**: The multi-exposure images are combined with geometric and photometric alignment. This approach requires a special hardware configuration; however, it can acquire the HDR scene luminance with high accuracy.

## 3.2. Spherical Aerial Image Completion Using Sky Model

### 3.2.1 Overview

In spherical images captured from the sky, there may be areas that an OMS cannot capture, or parts of the background scenery occluded by the vehicle. Missing areas must be filled in for these images to be usable as a part of an environmental light map upon which virtual objects are rendered. While an additional camera can be used for such purposes, as described in Section 3.3, it is usually difficult

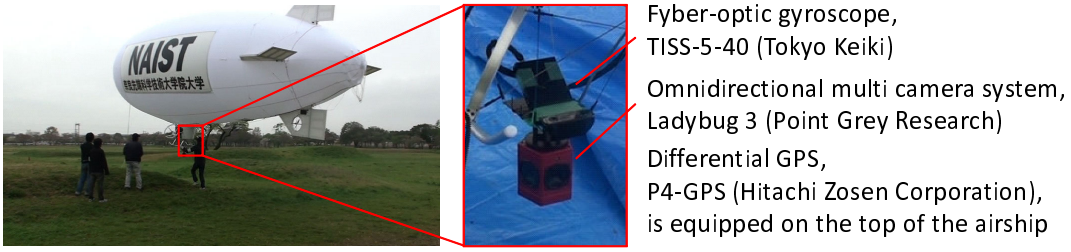
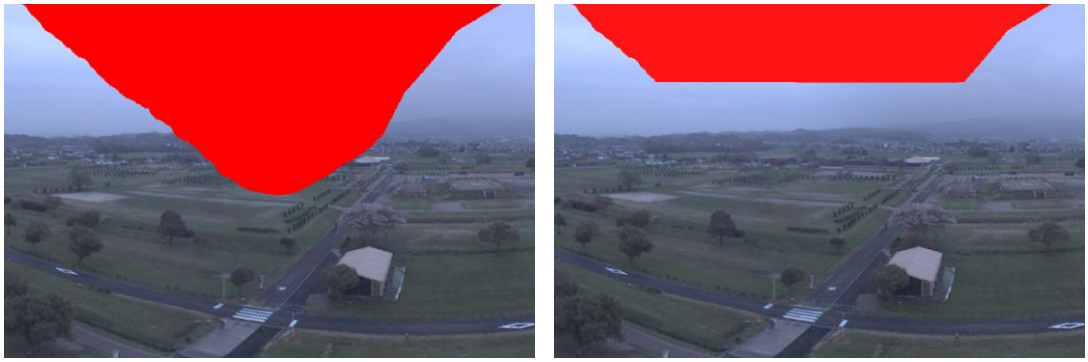



Figure 3.1. Unmanned airship and mounted sensors

to capture all directions in a single pass using aerial imaging. As in typical aerial imaging where the camera is mounted on the bottom of the capture vehicle, in our situation, missing areas occluded by the capture vehicle appear frequently in the sky. We overcome this problem using a sky model, which is a statistical luminance model of the sky for filling in missing areas. This is an easy way to acquire full spherical images from ordinary aerial imaging using an OMS. In this section, we describe a configuration for spherical aerial imaging using an OMS and an unmanned airship, followed by a completion method.

### 3.2.2 Aerial Imaging Using an OMS Mounted on an Aerial Vehicle

In our experimental settings, spherical images are captured using an OMS (Point Grey Research, Ladybug3) mounted on an unmanned airship, as shown in Figure 3.1. The airship was also equipped with a differential-GPS (Hitachi Zosen Corporation, P4-GPS), and a fiber-optic gyroscope (Tokyo Keiki, TISS-5-40). All sensors on the airship are connected to a laptop PC for storing the captured data. Figure 2.7 in Section 2.1.3 shows a panoramic image generated from perspective images acquired by six monocular cameras in an OMS using their intrinsic and extrinsic parameters [ISY03], with the missing area shown in the upper part. Six degrees-of-freedom (DoF) camera poses (3 DoF positions and 3 DoF orientations) are estimated from the multiple spherical images captured from the OMS by a structure-from-motion (SfM) approach. Using the camera pose information, the spherical images are aligned: the $i$-th image in the input sequence is spheri-



(a) Masked portion shows a missing area. A part of the missing area overhangs on ground scenery.

(b) Missing area on ground scenery is completed.

Figure 3.2. Completion of missing areas on ground scenery.

cally mapped, and the sphere is then rotated by $\mathbf{R}_i^{-1} = \mathbf{R}_i^{\mathrm{T}}$ using the estimated orientation $\mathbf{R}_i$ of the $i$-th image.

We assume that both the OMS and the vehicle are in the same position relative to each other and that the missing areas are manually identified. The position of a missing area does not change in the entire image set when the pose of the camera is relatively fixed to the capture vehicle. According to these assumptions, the completion process for an environmental map is as follows:

1. Completion of missing areas on ground scenery.
2. Completion based only on the intensity of the spherical images.
3. Completion of the remaining missing area using the All Sky Model [IKMN04].

### 3.2.3 Completion of Missing Areas on Ground Scenery

Although most parts of a missing area appear in the sky, the airship may also occlude a portion of the ground scenery, as shown in Figure 3.2(a). We suppose that there are no ground objects at a high elevation angle, and in this study, an area at a high elevation angle of over 5 degrees is defined as the sky area. The

other area is defined as the ground area. The intensities of the previous frame are used to filling-in a missing area in the ground section. We suppose that a change in the position of the OMS is sufficiently low between two consequent frames, and a missing area that appears is far away from the OMS, particularly for the ground area. A section similar to the missing area is searched from the previous frame using the Sum of Squared Differences (SSD), and the intensities of the corresponding pixels are copied to the pixels of the missing area. The ground area of the spherical video is filled-in as shown in Figure 3.2(b) by applying this process to the whole sequence.

Because this process assumes the input images as a sorted video sequence, the process should be skipped for unsorted multiple images. In this case, missing areas may remain on the resultant images.

### 3.2.4 Completion Based Only on the Intensity of Spherical Images

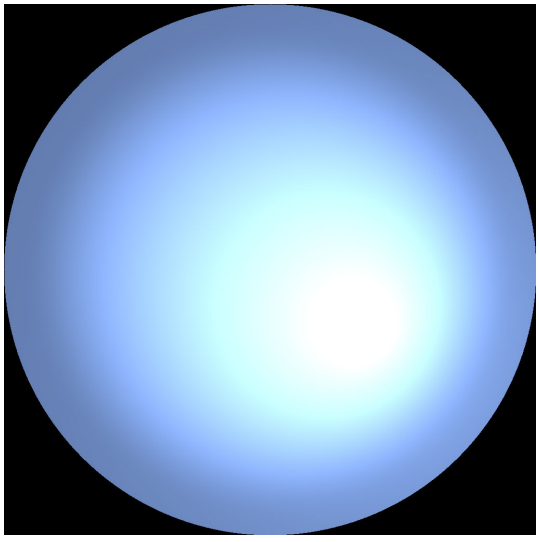
To estimate the intensity of the pixels in a missing area in the sky, we suppose that the illumination of the sky and clouds does not change during capturing the image sequence, and that the sky and clouds are infinite. Under these assumptions, the intensity of the sky is generalized, and one sky condition can be estimated in the entire image set because the intensity in the same direction does not change significantly while all images are captured. The aligned spherical images are first converted into *sky images* in equisolid angle projection, which have a uniform solid angle per pixel, as shown in Figure 3.3(a). The intensity $v_{uni}$ of a pixel in a direction in the unified sky image is estimated from the intensity $v_i$ of the pixel in the same direction in the $i$-th image of the input image set (see Figure 3.3(b)).

$$
\begin{aligned}
v_{uni} &= \begin{cases} \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i v_i & (\sum_i \alpha_i \neq 0) \\ undefined & (\sum_i \alpha_i = 0), \end{cases} \\
\alpha_i &= \begin{cases} 1 & (v_i \; is \; not \; in \; the \; missing \; area) \\ 0 & (v_i \; is \; in \; the \; missing \; area). \end{cases}
\end{aligned}
\tag{3.1}
$$

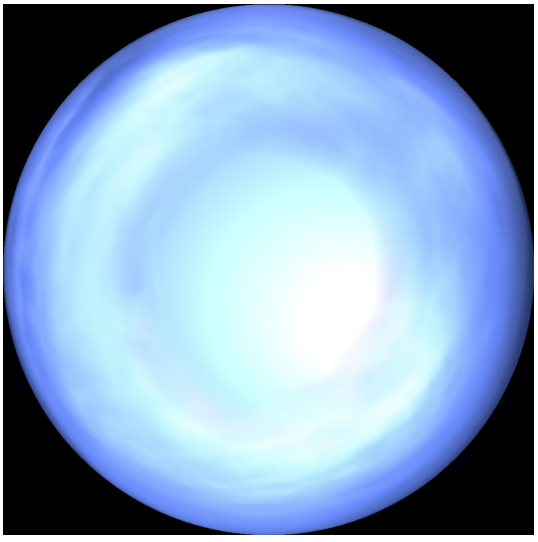
Because this process assumes that the illumination of the sky and clouds does not change, short sequences of only dozens or hundreds of frames are better when processing images of a drastically changing environment.

(a) A sky image in equisolid angle projection, where the missing area is masked.

(b) The completion results of the sky area based only on the intensity of the spherical video



(c) A sky image generated using All Sky Model [IKMN04]



(d) A completed sky image.

Figure 3.3. Completion of an environmental map.

### 3.2.5 Completion of Remaining Missing Area Using All Sky Model

The same part of a missing area can be occluded in the entire image set (i.e., $\sum_i \alpha_i = 0$); therefore, the intensity $v_{uni}$ cannot be determined in Equation (3.1). In the All Sky Model [IKMN04], the luminance (and radiance) distribution of the sky is statistically modeled, which can be used to complete the remaining missing area. The All Sky Model is known to be a good approximation of the sky distribution in all-weather scenarios using a sky index, which is a variable used for indicating the sky conditions. When generating a complete sky using the All Sky Model, the generated sky image is represented as an HDR image because the calculated intensities of the pixels can exceed the 8-bit value.

In this model, the sky luminance, $La(\gamma_s, \gamma)$, is defined as the product of the zenith luminance, $Lz(\gamma_s)$, and the relative sky luminance distribution, $L(\gamma_s, \gamma, Si)$. The relative sky luminance distribution varies depending on the weather. Thus, the All Sky Model [IKMN04] can be described as follows:

$$La(\gamma_s, \gamma, Si) = Lz(\gamma_s)L(\gamma_s, \gamma, Si), \tag{3.2}$$

$La(\gamma_s, \gamma, Si) : luminance\ of\ a\ sky\ element,$
$Lz(\gamma_s) : zenith\ luminance,$
$L(\gamma_s, \gamma, Si) : relative\ sky\ luminance\ distribution,$
$\gamma_s : direction\ (2\ DoF\ polar\ coordinates)\ of\ the\ sun,$
$\gamma : direction\ (2\ DoF\ polar\ coordinates)\ of\ the\ sky\ element, and$
$Si : sky\ index.$

Here, the sky index, $Si$, takes a value of $0.0 \leq Si \leq 2.0$ depending on the weather. A larger $Si$ indicates fine weather, and a lower value indicates a diffused (cloudy) sky. $Si$ is calculated using the amount of global solar radiation in the environment. Solar altitude $\gamma_s$ is calculated automatically using the acquired time and date of the images, as well as the location in latitude-longitude coordinates. Normally, the luminance of the sky, $La$, and the zenith luminance, $Lz$, are represented using a physical unit, which is an unknown value in our case. In this study, we used the

intensity of the pixels with linear gamma, which are proportional to the physical values.

The zenith luminance, $Lz$, which is required to calculate the intensity of the sky, is usually unknown because the zenith is occluded during an entire sequence. Equation (3.2) is modified to calculate $Lz$ from $La$ and $L$ as follows:

$$Lz = \frac{La(\gamma, Si)}{L(\gamma, Si)}. \tag{3.3}$$

Here, we treat $Si$ as a variable, because the global solar radiation, which is required to calculate $Si$, cannot be acquired due to existence of missing area. $\gamma_s$ is assumed to be a constant in our situation, which is calculated from the capturing time and date. The intensity of the pixels that are not in the missing areas can be used as $La$, but the $Si$ must still be estimated to calculate $L$. Once the $Si$ is estimated, $Lz$ can be calculated using the acquired $La$. The $Lz$ is ideally not varying, but when calculated using the real intensity it might include errors. We therefore estimate the $Si$ with the minimum variance of $Lz$ calculated from the intensity of each pixel in the sky images, which are generated by the method described in Section 3.2.4. Here, $V$ denotes a set of pixels fulfill two conditions: not belonging to missing area, and not saturated. If $\gamma_k$ denotes the direction of pixel $k$ in a sky image, the optimal sky index $Si_{opt}$ is estimated based on the minimum variance of $Lz$ as follows:

$$\begin{aligned} Si_{opt} &= \arg\min_{Si} \sum_{k \in V} \left( Lz_{acq}(\gamma_k, Si) - \bar{Lz}_{acq}(\gamma_k, Si) \right)^2, \qquad (3.4) \\ Lz_{acq}(\gamma_k, Si) &= \frac{La_{acq}(\gamma_k)}{L(\gamma_k, Si)}, \end{aligned}$$

where $Lz_{acq}$ denotes the calculated $Lz$ from the acquired real intensity $La_{acq}$, and $\bar{Lz}_{acq}$ is the average of $Lz_{acq}$ among all $\gamma_k$. Using Equation (3.4) is equivalent to calculating the $Si$ with the minimum error between the generated virtual skies, using every $Si$ and the real intensity, as illustrated in Figure 3.4. In our implementation, optimal $Si$ are searched from $Si = 0.0$ through $Si = 2.0$ by tiny increments (i.e., 0.1).

The intensity of the remaining missing area is calculated from Equation (3.2), using $Si_{opt}$ as the $Si$. The intensity may vary at the boundary between areas filled in using the method described in Section 3.2.4 and areas filled in using

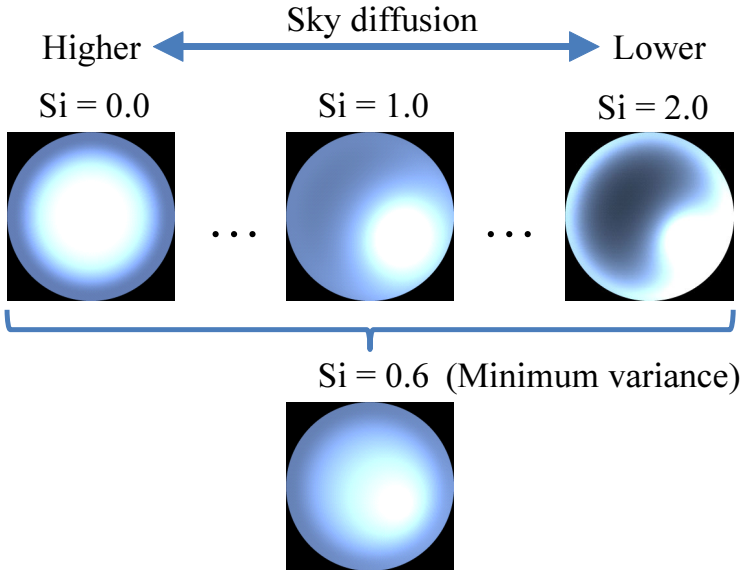
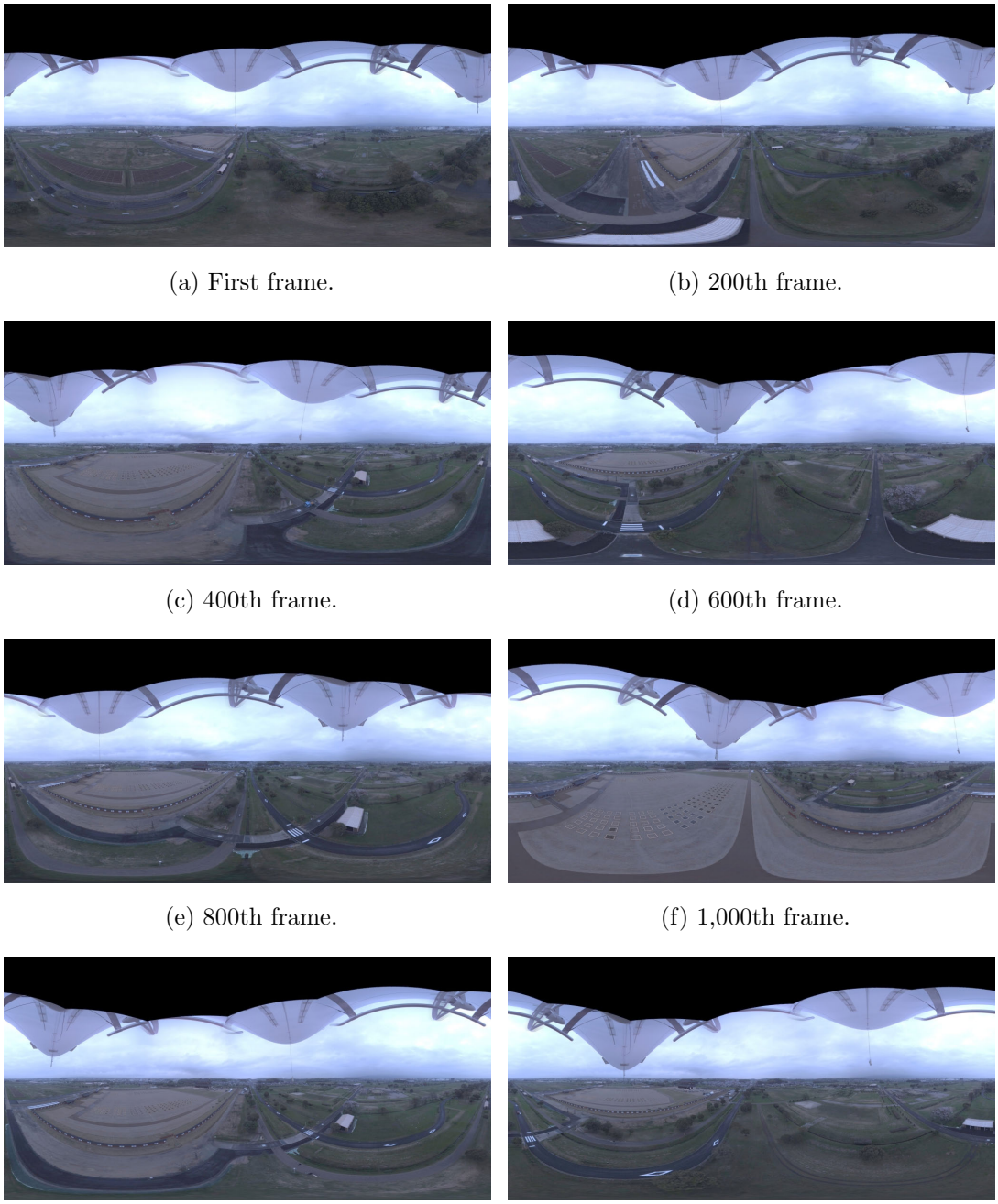



Figure 3.4. Estimation of $Si$ in the All Sky Model [IKMN04].

the method shown in Section 3.2.5; therefore, their boundary is alpha-blended into the estimated sky image. Figure 3.3(c) shows a sky image generated by the model, and Figure 3.3(d) shows an example of a complete sky image.

### 3.2.6 Experiment with Spherical Image Completion

To confirm that the spherical images are completed without unnatural artifacts, we conducted an experiment using a spherical image sequence captured by an OMS mounted on an unmanned airship. The spherical images were completed based on the method described in Section 3.2.2 using a sky model, where sky index $Si$, which indicates weather condition, for this experiment was estimated to be 0.6 (slightly overcast weather). The captured image sequence consisted of 1,900 frames, and the camera pose was successfully estimated for the whole sequence by a SfM approach using GPS measurements [YISY06]. Figure 3.5 shows frames sampled from the first 1,400 frames of the input image sequence. As in the completed images shown in Figure 3.6, full spherical images were successfully generated using the completion method without notable artifacts.



(a) First frame. (b) 200th frame.

(c) 400th frame. (d) 600th frame.

(e) 800th frame. (f) 1,000th frame.

(g) 1,200th frame. (h) 1,400th frame.

Figure 3.5. Spherical image sequence with missing areas, which were captured by OMS mounted on unmanned airship.

(a) First frame.

(b) 200th frame.

(c) 400th frame.

(d) 600th frame.

(e) 800th frame.

(f) 1,000th frame.

(g) 1,200th frame.

(h) 1,400th frame.

Figure 3.6. Full spherical image sequence completed using the sky model.

## 3.3. Full Spherical Aerial HDR Imaging Using Two OMSs

### 3.3.1 Overview

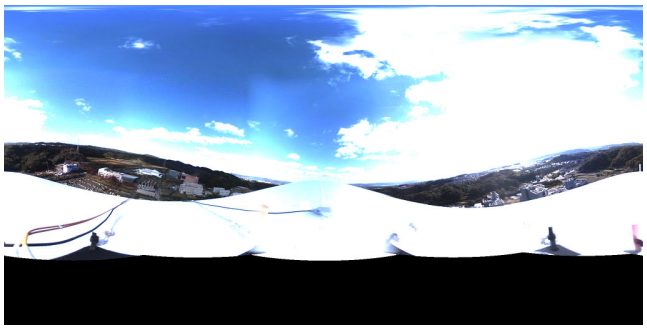
If a wealth of equipment was available for imaging, a new imaging technique could be employed, and the completion method described in Section 3.2 would not be necessary. This technique generates full spherical HDR images from multi-exposure images captured from two OMSs mounted on the top and bottom of an unmanned aircraft. That is, missing areas in a spherical image captured from the top (see Figure 3.7(a)) and bottom (see Figure 3.7(b)) of the vehicle are mutually completed, as shown in Figure 3.7(c), because the whole direction of the scene is acquired from at least one camera. Compared with the image completion technique, this approach is expected to acquire HDR scene luminance with higher accuracy.

As shown in Figure 3.8, this procedure is divided into the following three stages:

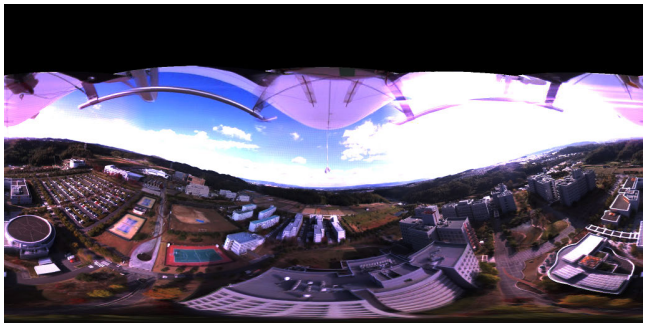
1. Multi-exposure aerial image capture from two OMSs.
   Multi-exposure images are captured using two OMSs mounted on the top and bottom of an unmanned aircraft. Since the camera on top of the vehicle mainly captures the sun and sky, it has neutral density (ND) filters attached. To capture high- and low-luminance images effectively, the exposure times are controlled automatically.

2. HDR image generation from multi-exposure images.
   HDR spherical images are generated from the multi-exposure images captured from the top and the bottom of the aircraft. Misalignments are corrected by estimating changes in the camera orientation over the sequence of captured images.

3. Composition of HDR images captured by the two cameras.
   Full spherical HDR images are synthesized from the HDR images generated during the second stage. Because the precise geometric relationship between images from the top and the bottom of the aircraft is difficult to fix (owing



(a) Image captured from top of the aircraft.



(b) Image captured from bottom of the aircraft.

(c) Full spherical image.

Figure 3.7. Full spherical HDR image generation using two OMSs.

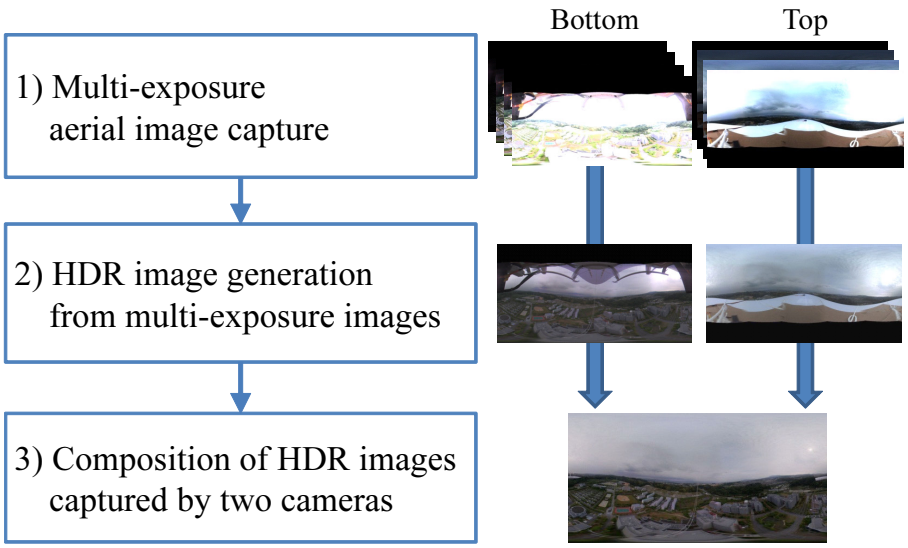



Figure 3.8. Overview of full spherical aerial HDR imaging.

to the slight deformability of the unmanned aircraft), we align the images by estimating the relative rotation from one camera to the other for each frame. We also correct chromatic changes that may occur in images captured using ND filtering.

### 3.3.2 Multi-Exposure Aerial Image Capture from Two OMSs

#### Configuration of the aerial imaging system using two OMSs

We used a 12-m remote-controlled aircraft, which is the same as the one used in Section 3.2.2. As shown in Figure 3.9, two OMSs (Point Grey Research Ladybug2) were mounted on the top and bottom of the aircraft. The cameras were connected to a laptop PC for time-stamped storage. ND filters (Fujifilm Corp., ND 2.0) blocking all but 1% of the visible light were attached to the camera on top of the vehicle. Figures 3.7(a) and 3.7(b) show panoramic images with the limb darkening removed [ISY03]. Note that the amount of limb darkening differs between images captured with and without ND filters.

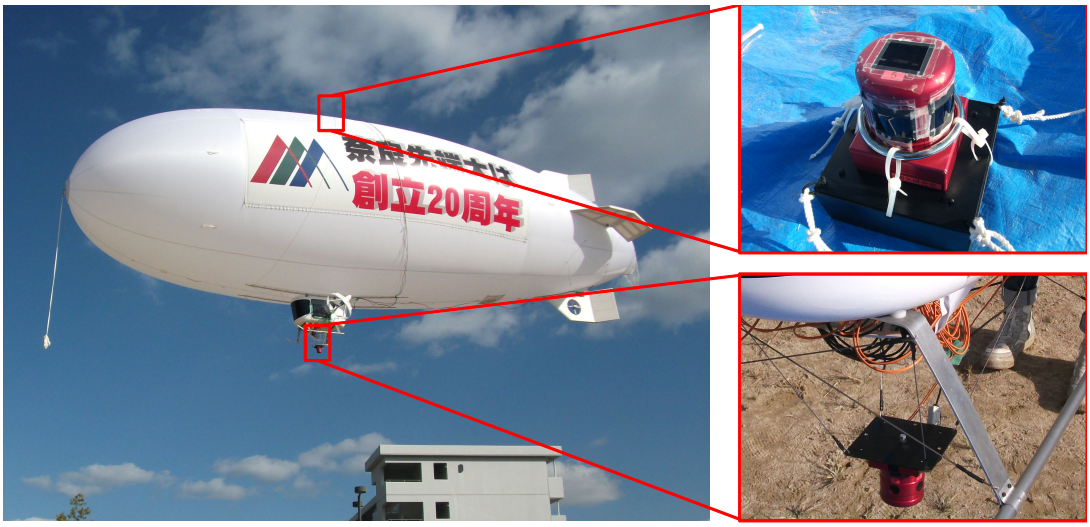



Figure 3.9. Unmanned aircraft and two OMSs. OMSs were mounted on both the top and bottom the aircraft. ND filters were attached to the top camera, and a GPS antenna was also mounted on the top of the aircraft.

### Automatic exposure control

The exposure times used to capture multi-exposure images are automatically determined from the intensity of the previously captured images. Many still cameras provide an auto-bracketing function for capturing multi-exposure images. They determine the proper exposure value using standard auto-exposure controls, and use fixed multiples of this value for the remaining exposure values [KUWS03]. Grossberg et al. [GN03] proposed a method for determining the unfixed exposure sets from the dynamic range of a scene. To determine a more appropriate exposure set, we use an HDR histogram of the previous multi-exposure image sequence to reduce the side effects of large quantization steps on human vision, such as pseudo-edges resulting from the use of too few multi-exposure images. Note that our approach optimizes the set of exposure times, whereas other modern studies have sought to minimize the signal-to-noise ratio (SNR) [HDF10]. Such minimization can also be integrated into our approach as a means for improving the appearance of HDR images, particularly those captured in dark environments.

New exposure times, $s_{new_1}, ..., s_{new_n}$, are determined from multi-exposure im-

ages captured using old exposures $s_{old_1}, ..., s_{old_n}$. In practice, for storing the exposure times, we set $n = 4$ as the number of exposures in our experiment to match the number of registers available in the Ladybug2 cameras.

The shortest exposure time, $s_{new_1}$, is the value at which the scene captured using $s_{old_1}$ can be done so without saturation, as follows:

$$s_{new_1} = \begin{cases} 0.5 s_{old_1} & (L_1 = MAX\_IN) \\ \dfrac{MAX\_IN + \theta_{sh}}{2L_1} s_{old_1} & (L_1 \leq \theta_{sh}) \\ s_{old_1} & (otherwise), \end{cases} \tag{3.5}$$

where $L_1$ denotes the maximum intensity in the image captured using $s_{old_1}$, and $MAX\_IN$ denotes the maximum representable intensity of the LDR image ($MAX\_IN = 255$ for 8-bit images). Further, $\theta_{sh}$ is a threshold that limits $s_{new_1}$ to $\theta_{sh} < L_1 < MAX\_IN$ when the response curve is regarded as linear.

When only a small number of multi-exposure images are acquired, side effects of the large quantization steps, such as pseudo-edges, can appear in the generated HDR images. Other exposure times, $s_{new_2}, ..., s_{new_n}$, are therefore determined to reduce such effects. To estimate the exposure times that have minimum quantization steps, an HDR histogram is calculated from the captured images, as illustrated in Figure 3.10. $H_i$ is the maximum intensity captured using the $i$-th shortest ($i \geq 2$) exposure time $s_{new_i}$ in the HDR histogram. Using $H_i$, $M_i$ is computed through

$$M_i = \sum_{k=\mathcal{H}}^{H_i} j_k, \tag{3.6}$$

where $j_k$ is the number of pixels whose intensity is $k$, and $\mathcal{H}$ is defined as

$$\mathcal{H} = \begin{cases} 0 & (i = n) \\ H_{i+1} & (otherwise). \end{cases} \tag{3.7}$$

The quantization step corresponding to $s_{new_i}$ in the HDR histogram can be calculated as follows:

$$\Delta_i = \frac{H_i}{MAX\_IN}. \tag{3.8}$$

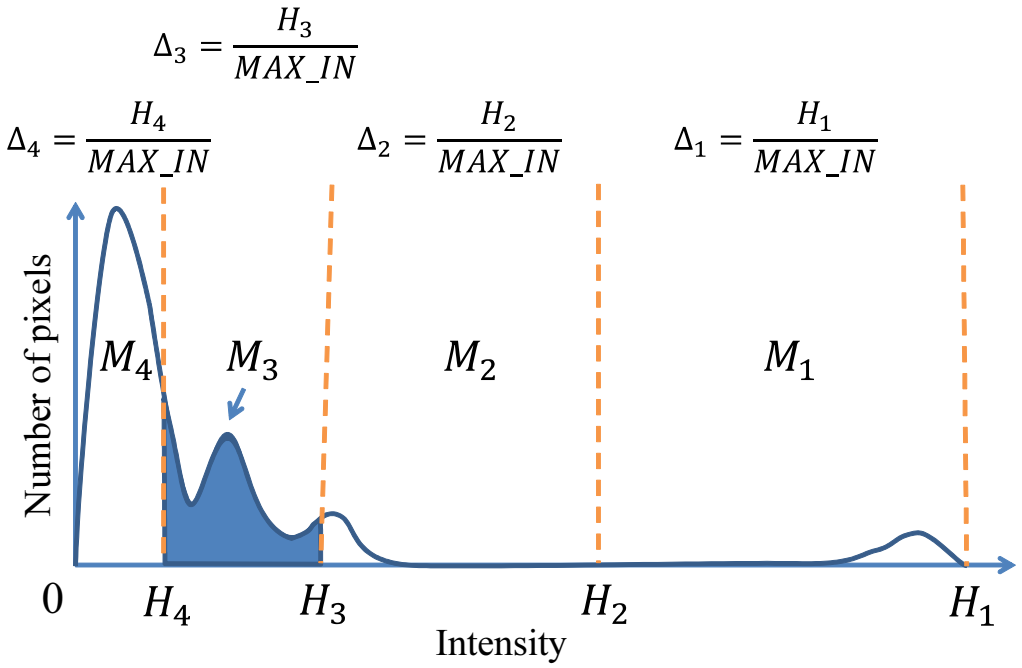


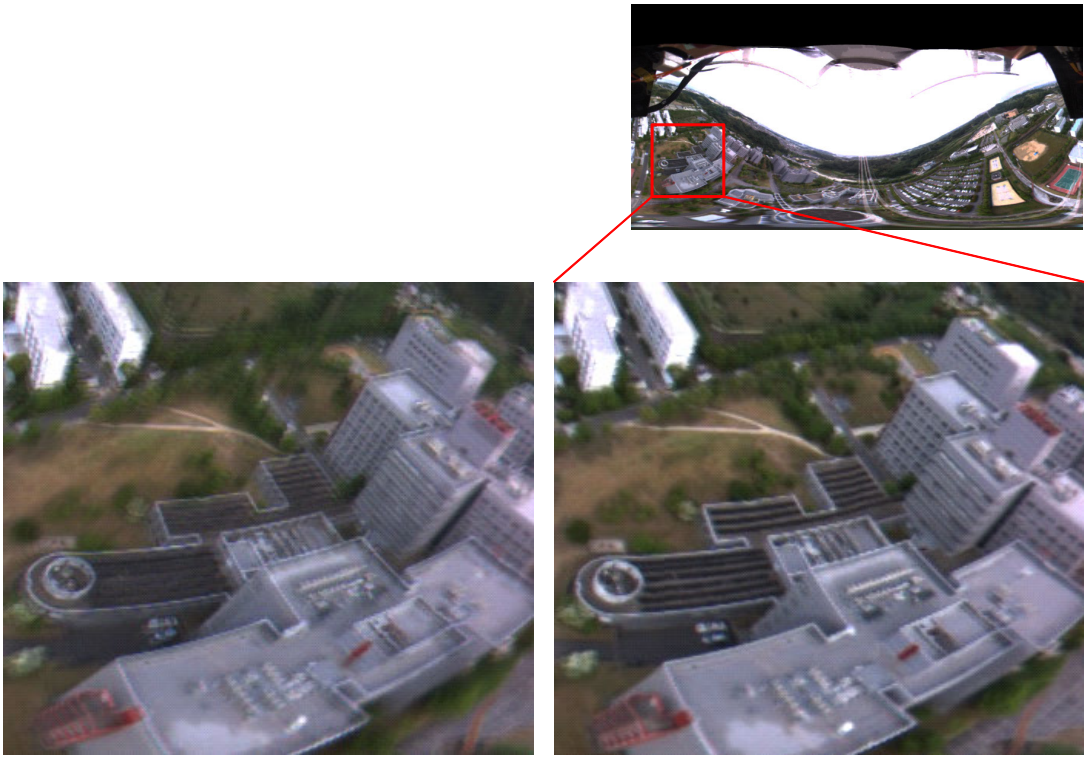

Figure 3.10. Illustrative example of HDR histogram for $n = 4$, where $\Delta_i$ denotes quantization steps and $H_i$ is the maximum intensity that can be captured using exposure time, $s_{new_i}$. Further, $M_3$ is the number of pixels in the filled area of the histogram, with the other $M_i$ determined in a similar fashion.

Finally, $s_{new_i}(i \geq 2)$ is determined by varying $s_{new_i}$ to minimize the energy function, $E_s$, where $E_s$ is the sum of the products of the quantization step $\Delta_i$ and the number of pixels corresponding to the exposure time $s_{new_i}$:

$$E_s = \sum_i (M_i \Delta_i). \tag{3.9}$$

$E_s$ is a non-linear function with multiple local minima. We apply a simple coarse-to-fine search technique to find a reasonable solution.

The above process for determining the exposure values is repeated every several seconds to allow for changes in the real-world lighting environment. Note that $s_{new_1}$ may not converge within a single cycle, and that several cycles are often needed.



(a) Without alignment. (b) With alignment.

Figure 3.11. HDR images with and without multi-exposure image alignment.

### 3.3.3 HDR Image Generation from Multi-Exposure Images

#### Alignment of multi-exposure images

Misalignments occur among multi-exposure images owing to changes in the position and orientation of the camera while capturing images at different exposure times. This causes blurring in the resulting HDR images, as illustrated in Figure 3.11(a). Our approach to resolving misalignments in aerial imaging is explained as follows.

For an aircraft moving at $5m/s$ and rotating at $30°/s$, the misalignment angles among multi-exposure images are calculated as shown in Table 3.1. When a

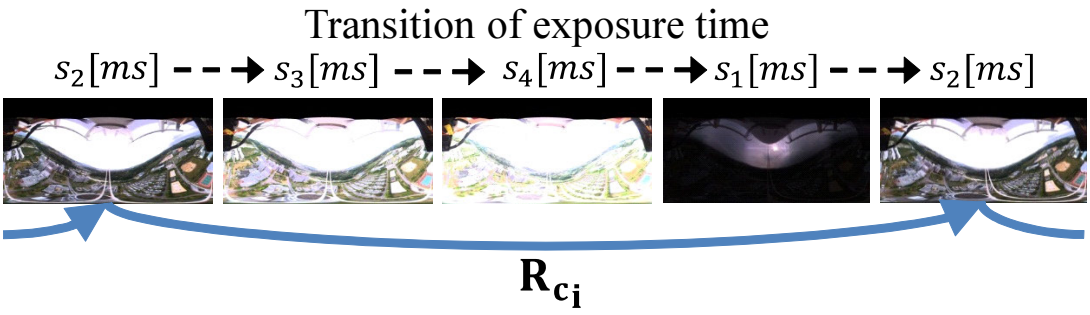
camera captures a scene from a great distance, which is a common situation in aerial imaging, the amount of misalignment is predominantly affected by changes in the camera orientation. We correct such misalignments by estimating the changes in camera orientation from the captured image sequence, as illustrated in Figure 3.11(b). Note that this correction method does not take a position change into consideration, and therefore ignores misalignments from fast-moving objects captured at short distances. Fortunately, local alignment methods, proposed for general multi-exposure imaging, can be applied when significant camera motion occurs.

Figure 3.12 shows the different steps of our multi-exposure image alignment process. In the first step, two images with the same exposure are selected, and the camera rotation between them is estimated, as shown in Figure 3.12(a). Among the multi-exposure images, those having the fewest saturated pixels (with an intensity of 255) and underexposed pixels (with an intensity of less than 16 in our implementation) are selected. The corresponding points $(\mathbf{p}_m, \mathbf{q}_m)$ between the selected images are determined using the KLT tracker [ST94] and projected onto a unit sphere. The parameters for rotation $\mathbf{Rc}_i$ are estimated by nonlinearly minimizing the energy function $E_l$, which is defined as the sum of the squared Euclidean distances between the projected corresponding points $|\mathbf{p}_m, \mathbf{q}_m|$:

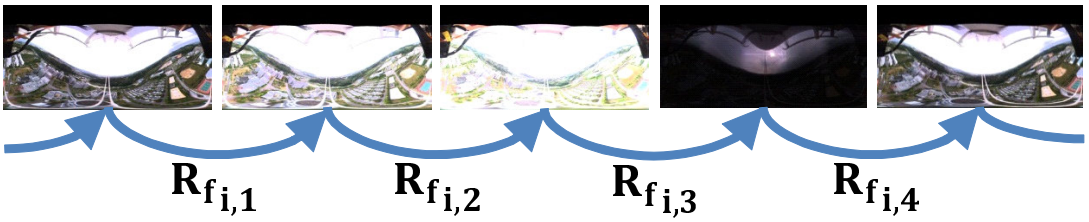
$$E_l = \sum_m |\mathbf{p}_m, \mathbf{q}_m|^2. \tag{3.10}$$

Table 3.1. Amount of misalignment owing to changes in position and orientation of the camera: one cycle represents the time taken to capture a single multi-exposure image set (0.25 s).

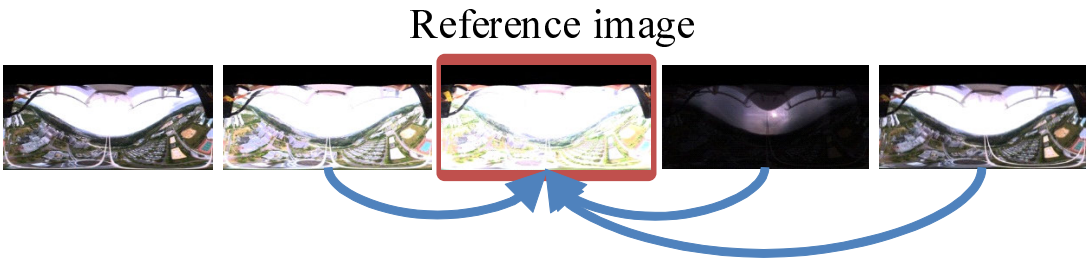
| | Distance to object | Misalignment per a cycle |
|---|---|---|
| Change of position | 50 m | 1.43° |
| | 100 m | 0.726° |
| | 500 m | 0.143° |
| Orientation | n.a. | 7.5° |




(a) Estimating rotation between images that have the same exposure.




(b) Calculating rotation between neighboring images using spherical linear interpolation.




(c) Calculating rotation to reference image from other images.

Figure 3.12. Process for aligning multi-exposure images.

Note that RANSAC [FB81] is used to reduce errors from mismatches.

Next, as shown in Figure 3.12(b), the rotation parameters between each two adjacent images $\mathbf{Rf}_{(i,j)}$ are calculated by interpolating $\mathbf{Rc}_i$ using a spherical linear interpolation.

The rotations of an arbitrary reference image from neighboring images are then calculated from $\mathbf{Rf}_{(i,j)}$, as shown in Figure 3.12(c). Multi-exposure images are aligned with the reference image through a transformation using the obtained rotation parameters. To align multi-exposure images for an entire video sequence, $\mathbf{Rf}_{(i,j)}$ must first be calculated for the entire sequence, after which each frame in the sequence can be treated as a reference image.

### HDR image generation

To compose the intensity of the multi-exposure LDR images, we use the HDR imaging method proposed by Debevec et al. [DM97]. In applying this method, it is necessary to consider the light attenuation owing to the ND filters attached to the camera on top of the aircraft. If the response curve is linear, or linearized in advance, the intensities in the HDR image, $I_h$, can be calculated using the LDR intensity $I_l$, the nominal light transmittance $\eta$ of the ND filter, and the exposure time $t$[s] of the LDR image:

$$I_h = \begin{cases} \kappa \frac{I_l}{t} & \text{(without ND filters)}, \\ \kappa \frac{I_l}{\eta t} & \text{(with ND filters)}. \end{cases} \tag{3.11}$$

Although the scale factor $\kappa$ is meaningless when only the relative pixel intensities are needed, it can be determined and used to calculate the absolute radiance values [W/sr/m$^2$]. The HDR intensities calculated from each set of multi-exposure images are composed in accordance with the method proposed described in [DM97].

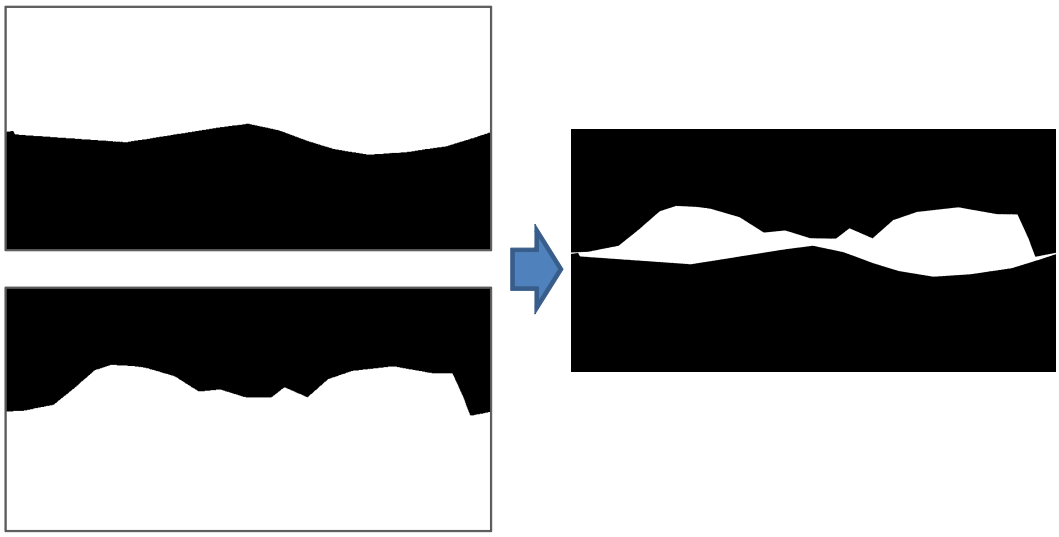


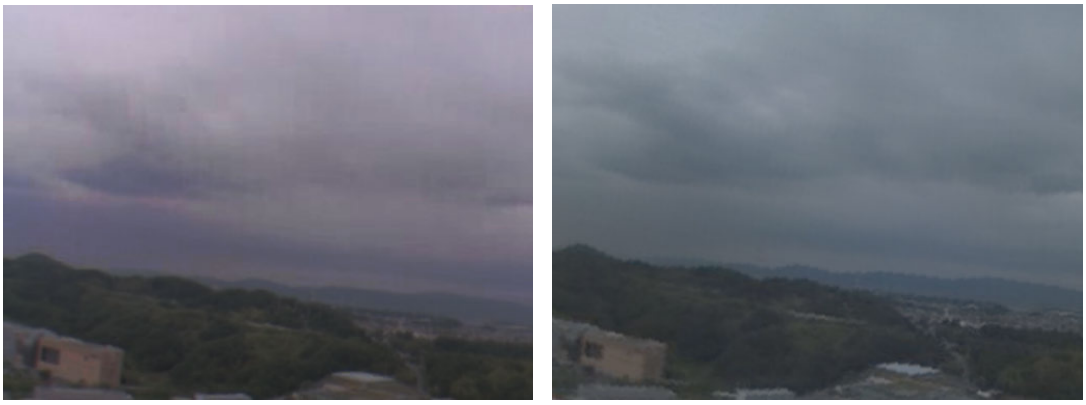
Figure 3.13. Masks used for alignment of images from the two cameras. Top left: Black pixels indicate the missing area from the top of the aircraft. Bottom left: Black pixels indicate the missing area from the bottom of the aircraft. Right: White pixels indicate an overlapping area.

### 3.3.4 Composition of HDR Images Captured by Two Cameras

#### Alignment between the two cameras

The HDR images captured from the top and bottom of the aircraft are aligned by estimating the relative rotation between the cameras. To estimate the parameters of this rotation, missing areas in the images are manually masked in advance, as shown on the left side of Figure 3.13. Note that only one mask image is required for an entire sequence captured by a given camera, as the missing area for spherical images does not significantly change. The area marked by white pixels on the right side of Figure 3.13 indicates the overlapping area where the scene has been captured by both cameras without occlusions. This area is defined as a conjunction of the negation of missing areas in the paired images.

The corresponding points are detected by applying the KLT tracker [ST94] to



(a) Without ND filter (image captured from bottom of airship). Average intensity: $(R, G, B) = (111.5, 108.3, 119.7)$.

(b) With ND filter (image captured from top of airship). Average intensity: $(R, G, B) = (95.1, 103.5, 111.9)$.

Figure 3.14. Example of chromatic change due to ND filters. A pair of HDR images converted to the same exposure is shown in close-up, at region where the two images overlap.

feature points in the overlapping area of the image captured from the bottom of the aircraft. The rotation parameters between the cameras from top to bottom are determined using the method described in Section 3.3.3

### Correction of chromatic change due to ND filters

Although ND filters are designed to transmit all wavelengths of light equally, they do cause a certain amount of chromatic changes [STJ+04]. The ND filters used in our experiment transmit less red light, as shown in Figure 3.14. Such changes are corrected by estimating the linear intensity transformation parameters from the intensities in the overlapping areas of the images from the two cameras. The RGB values $(R_{top}(\mathbf{x}), G_{top}(\mathbf{x}), B_{top}(\mathbf{x}))$ for pixel $\mathbf{x}$ in the image captured from

atop the aircraft are converted using the following linear transformations:

$$R'_{top}(\mathbf{x}) = \beta_r R_{top}(\mathbf{x}), \tag{3.12}$$

$$G'_{top}(\mathbf{x}) = \beta_g G_{top}(\mathbf{x}), \tag{3.13}$$

$$B'_{top}(\mathbf{x}) = \beta_b B_{top}(\mathbf{x}), \tag{3.14}$$

where $\beta_r$ is estimated by

$$\beta_r = \frac{\Sigma_{\mathbf{x}\in A}\frac{R_{bot}(\mathbf{x})}{R_{top}(\mathbf{x})}}{N_A}, \tag{3.15}$$

where $(R_{bot}(\mathbf{x}), G_{bot}(\mathbf{x}), B_{bot}(\mathbf{x}))$ denote the RGB values of the corresponding pixel $\mathbf{x}$ in the image captured from the bottom of the aircraft, $A$ denotes the overlapping area, and $N_A$ indicates the number of pixels belonging to $A$. In addition, $\beta_g$ and $\beta_b$ are estimated in a same fashion.

#### Combination of corrected HDR images

Full spherical HDR images are generated by combining paired HDR images that have been realigned and chromatically corrected, as described above. The intensities in the overlapping area of the full spherical image are determined by alpha blending the two images. Based on the intensities, $I_{top}$ and $I_{bot}$, of a pixel in the overlapping area of the two images, the intensity, $I_{full}$, of the corresponding pixel of the full spherical image can be calculated using

$$I_{full} = \alpha I_{bot} + (1-\alpha) I_{top}, \tag{3.16}$$

where $\alpha$ varies linearly between zero at the upper boundary of the overlapping area and unity at the lower boundary.

### 3.3.5 Experiment with Spherical Image Generation Using Two Cameras

#### Generation of still image

To confirm that the HDR image generated using our proposed method reflects the real environment with reasonable fidelity, we conducted an experiment in

which we generated a full spherical HDR image from still images captured using an unmanned aircraft. The aircraft was flown at 3 m/s at an altitude of 130 m while capturing the multi-exposure images. $MAX\_IN = 255$ and $\theta_{sh} = 192$ were used for this experimental environment.

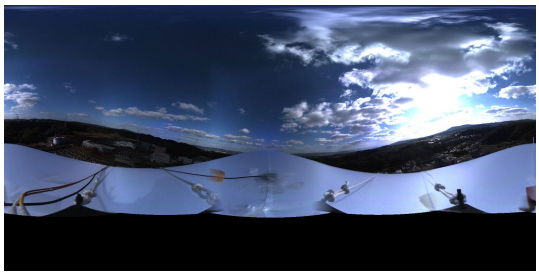
The captured multi-exposure images and corresponding exposure times are shown in Figure 3.15. Figures 3.7(a) and 3.7(b) show HDR images from the top and bottom of the aircraft, respectively. The full spherical HDR image composed from the images in Figures 3.7(a) and 3.7(b) is shown in Figure 3.7(c). Note that the two realigned and chromatically corrected images were successfully combined, leaving no obvious artifacts. Figure 3.16 shows the full spherical images whose intensity was cropped from the HDR image using the exposures equal to those in Figures 3.15(e) and 3.15(h). From this, we can confirm that the full spherical HDR image was satisfactorily combined from the captured multi-exposure images. The full spherical HDR image can also be visualized using typical tone-mapping methods such as the one in [RSSF02], as shown in Figure 3.17. Such tone-mapped images are suitable for immersive panoramas, which provide users with a sense of looking around a location, and the ability to see textures of various radiances.

**Generation of video sequence**

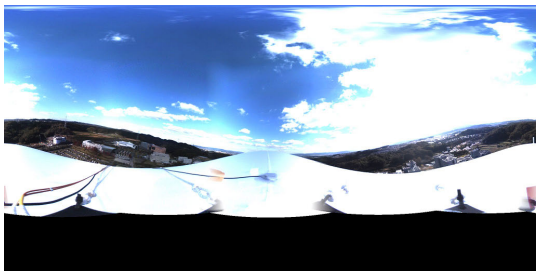
To generate full spherical HDR videos for fly-through applications, we applied our approach to a video sequence consisting of 500 frames. The aircraft was flown at a speed of 5 to 8 m/s and an altitude of 130 m while capturing the multi-exposure images. The frame rate of the generated HDR video was 16 fps, which is the same as that of the multi-exposure source video. The processing time was approximately 10 s per frame on a desktop PC with a Core i7-2600 CPU (3.40 GHz, 4 Cores). The position and orientation of the camera on the bottom of the aircraft were estimated using a SfM technique for OMS without GPS measurements [SIY04]. Figure 3.18 shows samples of the full spherical frames that are aligned using camera pose information by same manner as described in Section 3.2.2. Note that the position and orientation of the video frames were successfully estimated. This information can be used for a geometric registration between real and virtual objects in augmented immersive panoramas.

(a) Top: 0.1 ms (with ND filters).

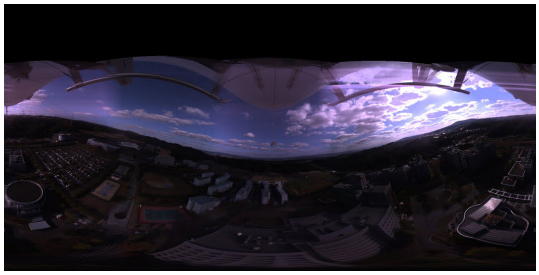
(b) Top: 4.1 ms (with ND filters).



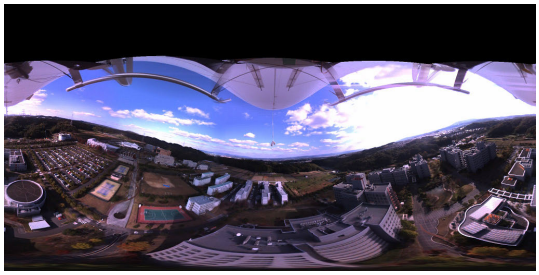
(c) Top: 14.4 ms (with ND filters).



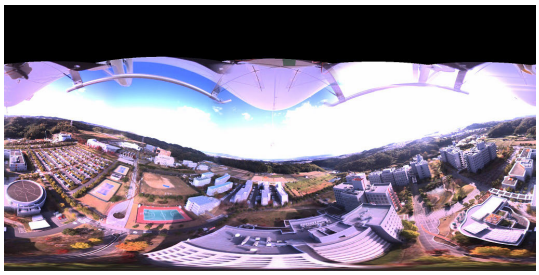
(d) Top: 43.3 ms (with ND filters).



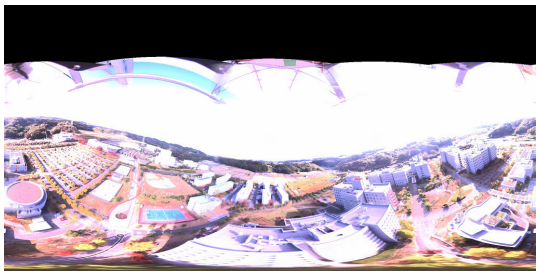
(e) Bottom: 0.1 ms.



(f) Bottom: 0.4 ms.

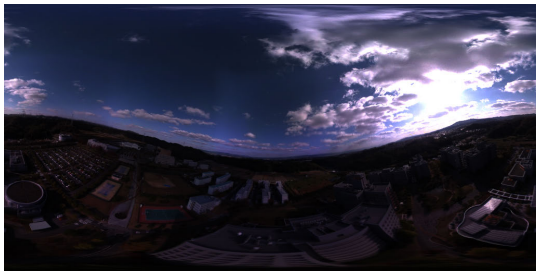


(g) Bottom: 1.1 ms.



(h) Bottom: 2.4 ms.

Figure 3.15. Captured multi-exposure images and corresponding exposure times.

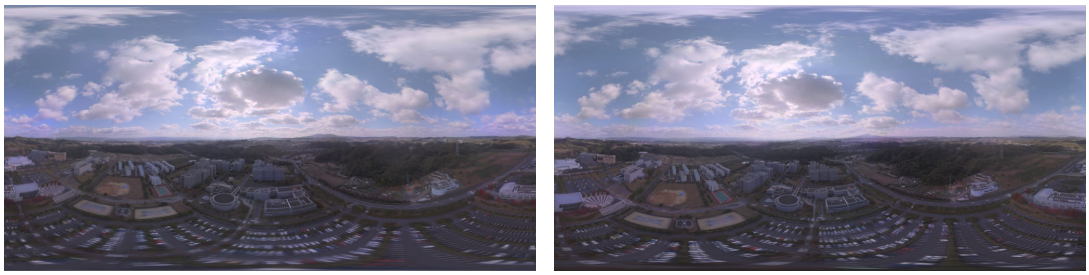


(a) Exposure for the image in Figure 3.15(e).

(b) Exposure for the image in Figure 3.15(h).

Figure 3.16. Full spherical images with cropped intensity using exposures of the multi-exposure images.

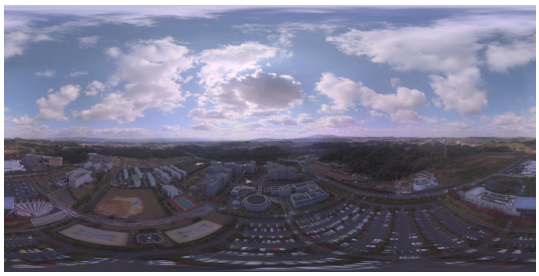
Figure 3.17. Full spherical image tone-mapped in accordance with the method by [RSSF02].



(a) First frame.



(b) 100th frame.



(c) 200th frame.

(d) 300th frame.

(e) 400th frame.

(f) 500th frame.

Figure 3.18. Full spherical tone-mapped images generated from video frames. The images were aligned using estimated relative camera orientation.

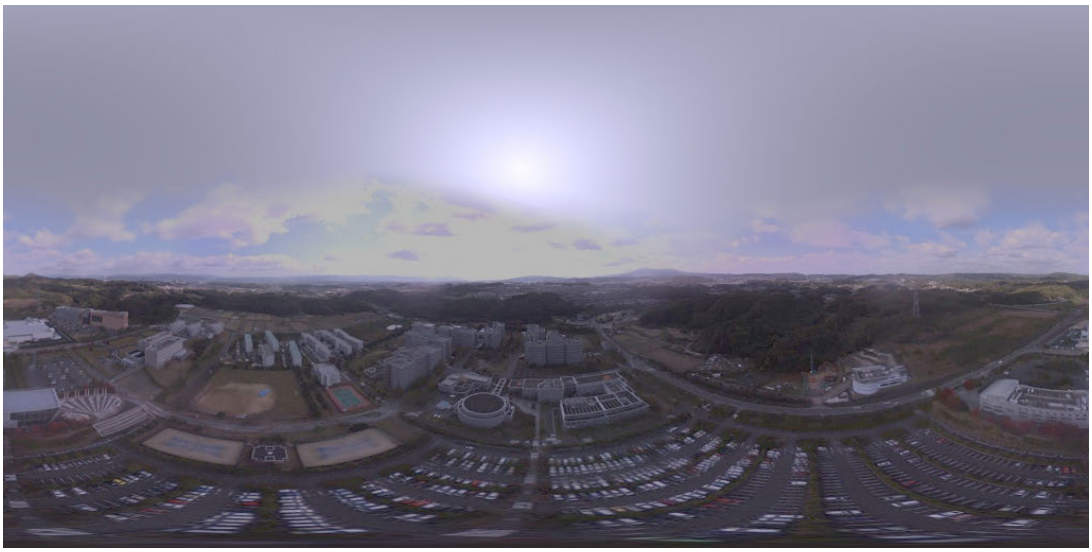


Figure 3.19. Completion of missing area based on spherical image completion using sky model.

## 3.4. Discussions

### 3.4.1 Comparison of Two Approaches

We compared the two approaches for acquiring full spherical aerial HDR images. Figure 3.19 shows the same frame as Figure 3.18(b) but with the missing area completed using the sky model. Note that, from a subjective viewpoint, the estimated *smooth* textures in Figure 3.19 may seem unnatural and decrease the immersive value of the image.

To investigate the effects when using full spherical images for a photorealistic virtual object superimposition using IBL, we compared the estimated intensity shown in Figure 3.19 with the acquired intensity. The average intensity of the sky estimated using the sky model was quite different from the intensity captured by the camera, $(1 : 0.218)$, as illustrated in Figures 3.18(b) and 3.19. When the image in Figure 3.19 is used for IBL, virtual objects with Lambertian surfaces become 0.218-times brighter than using the image in Figure 3.18(b). Furthermore, in Figure 3.18(b), the highest intensity (which affects the appearance of cast shadows) was ten-times larger than the intensity in Figure 3.19. These results indicate

that the approach described in Section 3.3, which employs an additional skyward camera for capturing the actual intensity of a missing area, produces better IBL rendering than current approaches using model-based sky area completion.

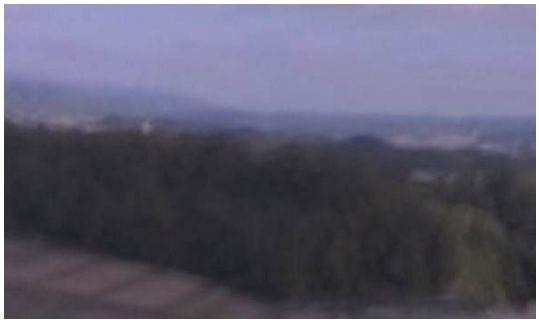
### 3.4.2 Auto-Exposure Control

We compared the auto-exposure control method described in Section 3.3.2 with ordinary controls for HDR imaging. Most control methods, including that used by [KUWS03], first determine the base exposure automatically, and then calculate the remaining exposure values at fixed stops from the base exposure. To avoid saturating the scene, we used the shortest exposure time, $s_{c_1}$[ms], as the base exposure for this experiment. The other exposure times, $s_{c_2}, ...s_{c_4}$, were then calculated for every two stops, meaning that $s_{c_i}$ was set to $2^2 = 4$ times longer than $s_{c_{i-1}}$ (i.e., fixed multiples). Note that advancing two stops for multi-exposure imaging is the default commonly used in commercial auto-bracketing cameras.

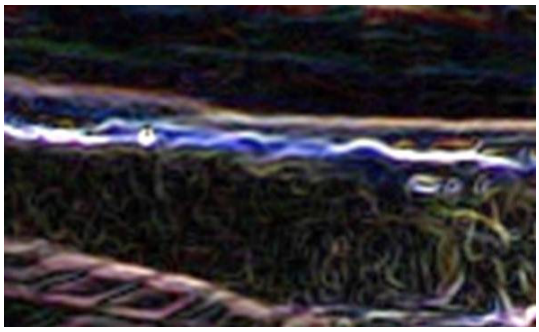
Table 3.2 shows the calculated maximum intensity and quantization steps for each approach. The maximum intensity of each multi-exposure image is a value

Table 3.2. Quantization steps of the HDR image using the approach described in Section 3.3.2 and the conventional approach based on fixed multiples of the shortest exposure time.

| | Exposure time[ms] | Acquirable max. intensity | Quantization step |
|---|---|---|---|
| Proposed | 0.1 | 1000 | 3.92 |
| | 4.1 | 24.4 | 0.10 |
| | 14.4 | 6.94 | 0.03 |
| | 43.3 | 2.31 | 0.009 |
| Fixed multiples | 0.1 | 1000 | 3.92 |
| | 0.4 | 250 | 0.98 |
| | 1.6 | 62.5 | 0.25 |
| | 6.4 | 15.6 | 0.06 |

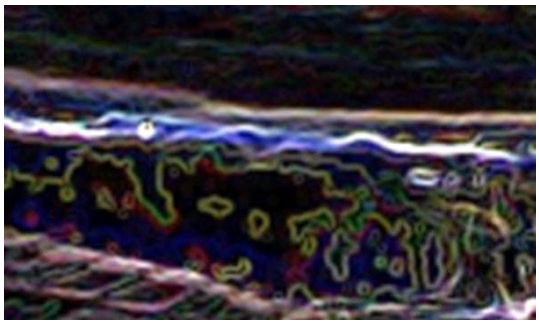


(a) Close-up of a generated HDR image.



(b) Gradient of the image in Figure 3.20(a) using Sobel filter.

(c) Image re-quantized using fixed-multiple exposures.



(d) Gradient of the image in Figure 3.20(c) by Sobel filter.

Figure 3.20. Visualization of negative effect of large quantization steps.

relative to the maximum intensity in the scene, which is defined as 1000. The quantization steps of the acquired multi-exposure images (8-bit) were converted into the HDR intensity by dividing the maximum intensity by 255 in the same way as $\Delta$ in Section 3.3.2. A smaller value is better here, as it help prevent visual artifacts such as pseudo-edges. A comparison of the longest exposure time indicates that the quantization step of the conventional method was 6.7-times larger than that of our proposed approach. This difference can particularly affect areas of low intensity, such as most ground features. Figure 3.20 shows close-ups of a generated HDR image of the same frame shown in Figure 3.17. Figure 3.20(c) shows an artificially re-quantized image generated from this image using the exposure times of a conventional approach, $s_{c_i} (1 \leq i \leq 4)$. Figures 3.20(b) and

3.20(d) show the gradient images of the frames in Figures 3.20(a) and 3.20(c), respectively. Pseudo-edges appear in the image in Figure 3.20(d), mainly in low intensity areas. This indicates that our proposed approach can prevent visual artifacts produced when using a conventional approach.

## 3.5. Summary

This chapter described two approaches for acquiring full spherical aerial images suitable for photorealistic MR-world generation. The first approach, which was described in Section 3.2, only requires ordinary spherical aerial imaging equipment consisting of an aerial vehicle and an OMS mounted on the bottom of the vehicle. Spherical images captured from the equipment include missing areas, and are therefore completed using the All Sky Model [IKMN04], which is a statistical model of the sky luminance and radiance. The second approach requires special hardware configurations, that is, two OMSs mounted on the top and bottom of an aerial vehicle, as detailed in Section 3.3. The two OMSs cover the whole direction at a particular viewpoint; therefore, full spherical images are generated by completing mutual missing areas appearing in the spherical images captured by each OMS. In addition, to generate HDR images, the two OMSs each capture multi-exposure images.

Although both approaches generate HDR results, the experiments demonstrated that the approach using two OMSs generated a notably higher intensity compared with the spherical image completion using a sky model. This result indicates that the additional skyward camera largely improves the accurate acquisition of real-world illumination, which can be used for virtual-object superimposition using IBL. When we employ the results of model-based sky completion for IBL, it may be necessary to manually edit the illumination information.

# Chapter 4

# Photorealistic Rendering for MR-World Exploration

## 4.1. Overview

In some applications that allow an interactive virtual exploration of a virtualized real-world, virtual objects can be superimposed for a visualization of disaster areas or non-existent buildings [XBF+09, OKY10]. As discussed in Chapter 1, the virtual objects are rendered frame by frame in applications that allow users to change their viewpoint. Unlike the previous real-time rendering used in AR, this study proposes a framework that combines the offline rendering of virtual objects and free-viewpoint image generation to take advantage of the high quality of offline rendering without the computational cost of online CG rendering, i.e., it incurs only the cost of online computations for free-viewpoint image generation. In addition, the generation of structured viewpoints (e.g., at every grid point) reduces the computational costs required in an online process. The advantages of our framework are as follows:

- High-quality superimposition of virtual objects by combining pre-rendering and free-viewpoint image generation.
- Reducing the computational cost in online free-viewpoint image generation using pre-generated structured viewpoints.

This chapter describes a practical implementation that provides a free-viewpoint using a 2D grid structure based on our framework, which superimposes virtual buildings onto real scenes virtualized. This application inputs full spherical aerial images acquired by the method described in Chapter 3. The virtual objects are rendered based on IBL using spherical images.

## 4.2. Free-Viewpoint Image Generation Framework for Photorealistic Superimposition of Virtual Objects in Real-World Virtualization

The proposed rendering framework inputs multiple spherical images capturing the real world, including but not limited to spherical aerial images acquired from the imaging techniques proposed in Chapter 3. The rendering framework can basically be employed for general situations using spherical images.

Our framework is divided into two parts: an offline process and an online process. The offline process inputs both real-world images captured at various positions, and 3D models of the virtual objects, and generates augmented scenes at numerous viewpoints. Depending on the target environment, the viewpoints are given as a multi-dimensional structure, such as on a line, grid, or cube, as illustrated in Figure 4.1. Real-world views from these structured viewpoints are first generated using a free-viewpoint image generation method using spherical images and reconstructed 3D shapes of the real-world objects. If the viewpoints are densely designated, the quality of online-generated images is expected to increase; however, large amount of data storage is required to store pre-generated images. The offline process can include manual operations, and therefore, in principle, cinema-quality superimposition can be achieved with time and effort.

The online process generates a perspective image from the user's location through real-time free-viewpoint image generation using offline-rendered textures. This process does not depend on the method used to estimate the position and orientation of the user's viewpoint, and does not designate a specific free-viewpoint image generation method. A simple guideline for selecting proper free-viewpoint image generation methods is described in Section 2.2.1. The structured view-

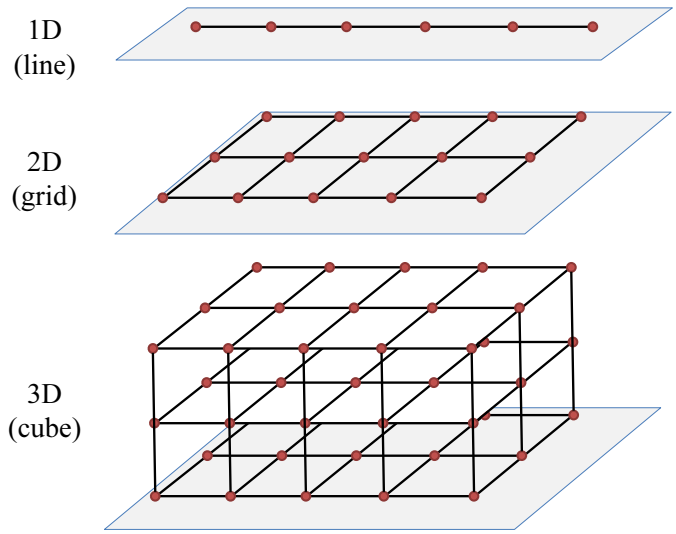



Figure 4.1. Structured viewpoints designated during offline process.

points, which are generated in the offline process, are basically expected to reduce the computational costs of the online process. In image-based or hybrid free-viewpoint image generation, the texture selection and blending processes refer to $m$ multi-viewpoint images, with respect to each of $n$ polygons of the 3D shape in physically-based methods, or $n$ pixels of the image in appearance-based methods. This process can include up to $O(m * n)$ calculations. The online process in the proposed framework realizes $O(1)$ calculations for texture selection and blending, which refers only to neighboring structured viewpoints, as illustrated in Figure 4.2. In principle, the dimension of user's location changeable in online can be differ from that of the structured viewpoints; i.e., 3D viewpoint change can be realized using one dimensional (1D) or 2D structured viewpoints. However, it should receive careful attention that image generation from a location where faraway from any structured viewpoints clearly leads degradation of the image quality.

As described above, the rendering framework consists of simple processes and is able to accommodate various methods. In the remainder of this chapter, the proposed framework is detailed using an example of our specific implementation.

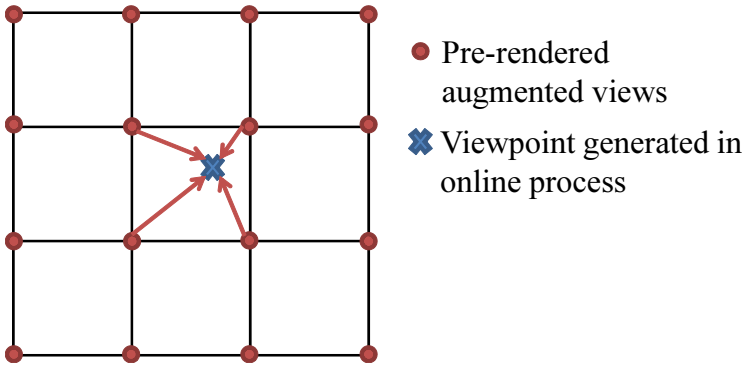



Figure 4.2. Online free-viewpoint image generation using only neighboring structured viewpoints (in case of 2D structured viewpoints.)

## 4.3. Fly-Through Application Using Full Spherical Aerial Images, 2D Grid Structure, and VDTM

This section describes an example of practical implementation based on the proposed rendering framework, using full spherical aerial images generated as described in Section 3. Note that, proposed framework itself does not restrict the input to the aerial images. The application is summarized as follows:

- **Intended use**: A virtual tourism application superimposing 3D models of historical buildings. The user's viewpoint is configured in the sky; it is intended to be used like virtual globe applications.
- **Input**: Full spherical aerial images generated as described in Section 3.
- **Dimension of structured viewpoint**: 2D grid.
- **Free-viewpoint image generation method**: VDTM [DTM96]. Because this application employs aerial views, it is expected that the 3D shapes of the real-world are far from the viewpoints. As discussed in Section 2.2.1, physically-based hybrid rendering approaches should be appropriate for our application. VDTM is one of the common physically-based hybrid rendering approaches.

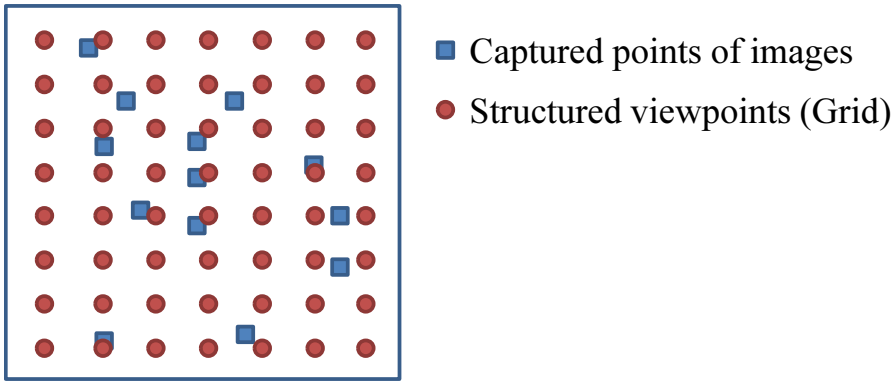



Figure 4.3. Pre-generation of real-scene images at grid points: unstructured captured viewpoints are re-sampled into a grid viewpoint structure using offline free-viewpoint image generation.

In our application, real-world views are first generated at every grid point, as illustrated in Figure 4.3, using a VDTM approach [DTM96]. Virtual objects are then photorealistically rendered for every structured viewpoint using real-world illumination without any missing areas of the illumination environments in the spherical images, which are generated using the completion method described in Section 3.2.

Augmented free-viewpoint images, whose viewpoint can be freely configured on a plane, are presented to users during the online process. The online free-viewpoint images are generated from structured augmented viewpoints, which are generated during the offline process. The computational cost of online free-viewpoint image generation is reduced using a simplified version of VDTM, where the textures of four neighboring structured viewpoints are blended into simplified 3D models using a bilinear weighting function (as illustrated in Figure 4.2). The online process requires only four (i.e., O(1)) weight calculations to blend textures, while the offline process needs $O(m * n)$ calculations with $m$ polygons and $n$ cameras.

### 4.3.1 Offline Process: Photorealistic Superimposition of Virtual Objects on Structured Viewpoints

During the offline process, the augmented images are generated at the structured viewpoints (at every grid point in our implementation) from the real-world images captured at hundreds of positions and from 3D models of the virtual objects. The flow of the offline process is as follows:

1. Generation of real-scene images for structured viewpoints
2. Rendering of virtual objects for structured viewpoints

Camera pose information and the 3D shapes of real-world objects are used to generate the views from the grid points. The views are constructed from the input images using VDTM, a hybrid free-viewpoint image generation of model- and image-based rendering [DTM96]. The virtual objects are superimposed with the IBL on the completed environmental maps of the structured views, which are generated for every grid point.

#### Generation of real-scene images for structured viewpoints

Six DoF camera pose and dense real-world 3D shapes are first reconstructed. We employed a modern vision-based SfM approach using captured spherical images and GPS measurements. VisualSFM [Wu13], which is an SfM application using multi-core bundle adjustment [WACS11], estimates the relative camera positions; however, it does not provide an absolute scale, which is required for geometric registration between real and virtual objects without adjusting their coordinates manually. For our situation, camera position information using an absolute scale was acquired from the GPS installed on our aerial vehicle, and a non-linear minimization of the SSD between the GPS measurement and the camera position estimated using SfM was performed. Some approaches combining SfM and GPS measurements using constrained/extended bundle adjustments [KTSY10, Lhu11] may also be suitable to our situation with slight improvements. CMPMVS [JP11], a state-of-the-art multi-view stereo (MVS) software, was used for further reconstruct dense 3D polygonal shapes of real-world objects. The spherical images were preliminarily converted into six perspective images (cube maps) to prepare

the input data for the SfM and MVS software, which uses perspective images as the input. The estimated camera pose is also used for the completion of missing areas in spherical images as described in Section 3.2. Note that the input images for this application are not prepared as a video sequence; a part of the completion process assuming video input, as described in Section 3.2.3, is not performed.

The reconstructed 3D models of large outdoor environments consist of millions of polygons, which causes a heavy amount of computations for free-viewpoint image generation. Our framework therefore generates densely packed structured viewpoints of the real world during the offline process to reduce the amount of computations of the online process. By generating dense arrangement viewpoints, the 3D models are expected to be greatly simplified during the online process because of the small disparity between neighboring pre-generated viewpoints. In our application, spherical views of the physical environment are pre-generated at every grid point on the designated plane, allowing changes in the user's viewpoint, as illustrated in Figure 4.3.

In the VDTM [DTM96], the images captured from multiple positions are projected and blended appropriately into each reconstructed polygon. A large hemispherical geometry, assumed to be infinite, is used as the background geometry for portions of the scene that are not reconstructed, such as the sky. The color, $Ip_j$, of the surface on the $j$-th polygon is determined by alpha-blending the texture color, $Ic_i$, projected from $i$-th camera, depending on the viewpoint to be generated. $Ip_j$ is calculated using the weight for blending $\omega_{off}(i)$, which is the reciprocal of the penalty function, $\omega_{ang}(i)$, as follows:

$$\begin{aligned} Ip_j &= \sum_i \frac{\omega_{off}(i) \cdot Ic_i}{\sum_i \omega_{off}}, \\ \omega_{off}(i) &= \frac{1}{\omega_{ang}(i) + \epsilon}. \end{aligned} \tag{4.1}$$

Here, $\omega_{ang}(i)$ denotes the distance between the $i$-th camera and the desired ray, which is defined as a vector from the viewpoint to the center of the polygon (see Figure 4.4). In addition, $\epsilon$ is a tiny value used to avoid dividing by zero. Note that our definition of the penalty $\omega_{ang}(i)$ is commonly used in selecting appropriate rays, such as in an image-based stereo-view generation approach based on light-ray selection [HKY10]. The visibility of the 3D shape from the $i$-th camera is

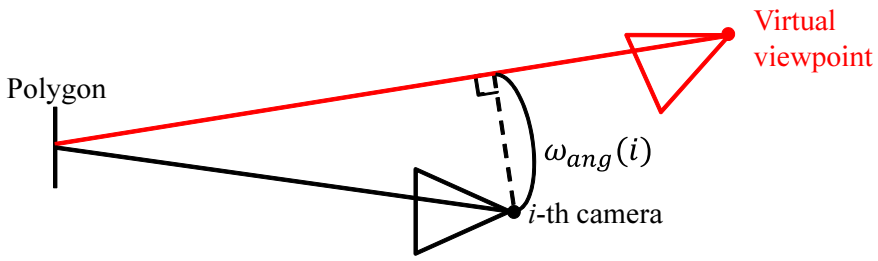



Figure 4.4. Penalty definitions of our offline VDTM.

calculated using the camera's depth maps as in a study on VDTM using the per-pixel visibility [PDG05].

The value of $\omega_{off}(i)$ varies depending on the positions of the polygons, i.e., this process requires $m$-times penalty calculations for the $i$-th camera when there are $m$ polygons. As the worst case, the calculation of the weight $\omega_{off}(i)$ requires O($m*n$) with $n$ cameras. Even if the calculation process is reduced using only the $k$-th best cameras, as in [BBM+01], or using a graphics processing unit (GPU) for parallelization, the rendering process with the binding and switching of thousands of textures generates a large amount of overhead for graphic pipelines.

### Rendering of virtual objects onto structured viewpoints

Virtual objects are rendered onto each structured real-world spherical view using a commercial GI-rendering engine, such as 3ds Max (Autodesk, Inc.) or Mental-ray (Mental Images GmbH.). Camera pose information, estimated through SfM, is used for geometric registration between the virtual objects and the real environment. The augmented scenes are generated through IBL and GI rendering, as in commercial movies, using complete environmental maps and dense 3D shapes of the real world to create ambient and occlusion effects. It should be noted that the illumination environment for virtual objects is manually edited from the original full spherical images by setting up additional lights. The allowance of manual editing is an important advantage of the proposed rendering framework.

Figure 4.5 shows an example of an augmented spherical image at a grid point. The real-world scene was generated using VDTM, as described in Section 4.3.1, and the CG models of the buildings were rendered offline.

Figure 4.5. Example of augmented spherical scene at a grid point.

### 4.3.2 Online Process: Free-Viewpoint Image Generation with Reduced Computational Cost

The online process generates planer perspective images from freely configured viewpoints in real-time, using augmented spherical images at the structured viewpoints and simplified 3D models of both real and virtual objects.

We enabled the free-viewpoint image generation on a 2D plane using the pre-generated structured viewpoints at the grid points. VDTM is improved so as to reduce the amount of online computations. During the online process, the views at four neighboring grid points are projected onto 3D surfaces and blended with bilinear weights, which were calculated based on the positions of the grid points and the viewpoint to be generated, as illustrated in Figure 4.2. Thus, Equation (4.1) is modified to use the bilinear weight, $\omega_{bilinear}(k)$, as follows:

$$Ipr_j = \sum_{k=1}^{4} \omega_{bilinear}(k) \cdot Icr_k, \tag{4.2}$$

where $Ipr_j$ denotes the surface color on the $j$-th polygon and $Icr_k$ is the color of the pixel in the $k$-th ($1 \leq k \leq 4$) neighboring view projected onto the surface. This process does not require $O(m * n)$ weight calculations because $\omega_{bilinear}(k)$

does not depend on the positions of the polygons, but only on the position of the viewpoint. It requires only four calculations for $\omega_{bilinear}(k)$ while generating an image.

In our application, 3D models of both real and virtual objects are combined and simplified down to 1.4% of the original numbers of polygons using a quadric-based mesh decimation method [GH97]. Although the simplification of the 3D models does not reduce the cost of calculating $\omega_{bilinear}(k)$, the rendering cost for the 3D models in a graphics pipeline is linearly reduced, depending on the number of polygons. The online process can be easily implemented on a GPU, such as using OpenGL Shading Language (GLSL), with a per-pixel visibility test using the depth maps at the structured viewpoints [PDG05].

## 4.4. Experiment of Application Using 2D Structure

To confirm that the application based on the proposed framework generates the appropriate scenes in a practical environment, augmented free-viewpoint scenes were generated on a 400m $\times$ 400m plane at an altitude of approximately 50 m from the ground. The application setting was based on the palace site of Heijo-kyo, an ancient capital city of Japan, where the original palaces no longer exist. The grid points, which were used for the positions of the structured viewpoints, were designated for every 20m $\times$ 20m area. The positions of the captured spherical images and the configured grid are shown in Figure 4.6. The input spherical images were captured at various altitudes, and the altitude of the grid plane was designated as the average altitude of the captured points. The superimposed virtual buildings of Heijo Palace, presented in Figure 4.7, were approximately 1 km in length from east to west and 1.3 km from north to south, and consisted of 4,255,650 polygons. The real-world environment was reconstructed using 3,290,880 polygons from 174 spherical images. During the online process, 3D models of both real and virtual worlds were combined and simplified into a model of 104,054 polygons (approximately 1.4% of the original 3D model), as shown in Figure 4.8.

Examples of augmented images, which were generated during the online process, are shown in Figure 4.9. The viewpoint of each image is shown in Figure 4.6.

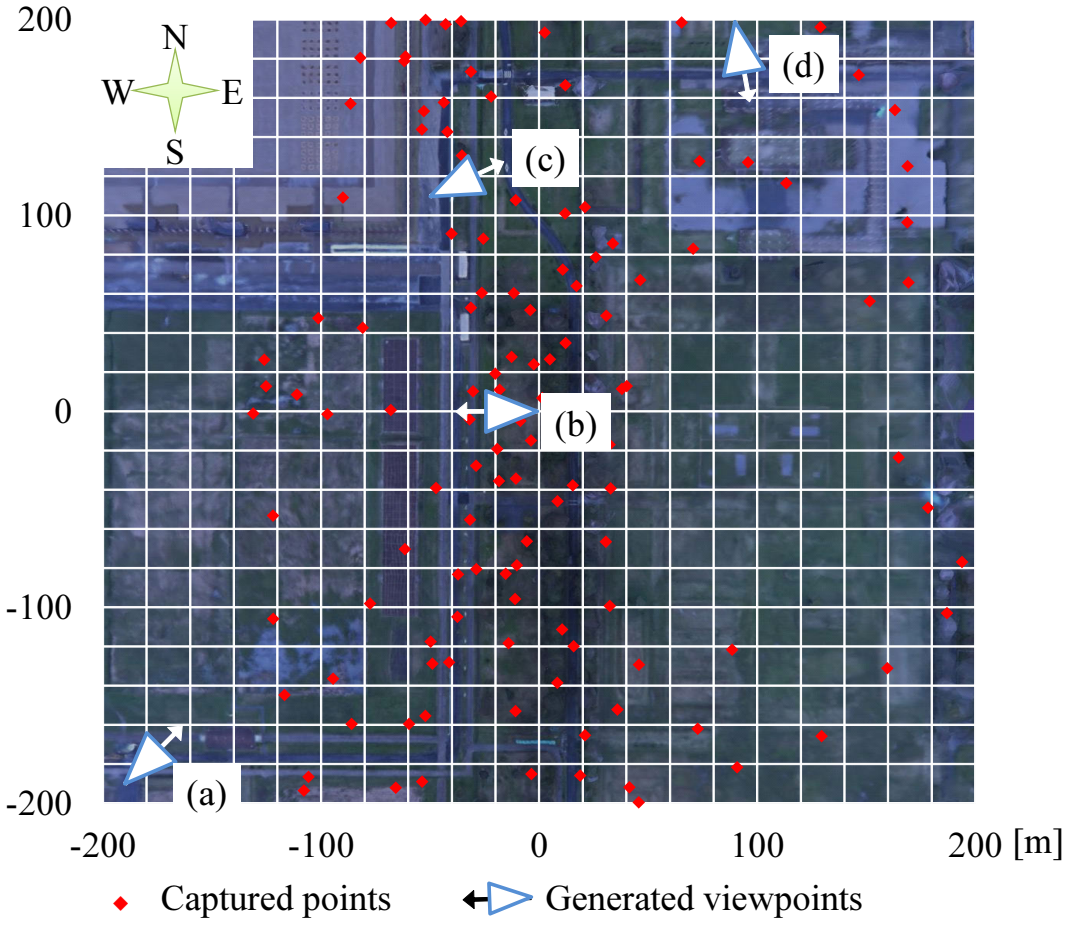



Figure 4.6. Captured points of spherical images, structured viewpoints, and online-generated viewpoints. Structured viewpoints were designated at every 20m $\times$ 20m grid point area.

The online process with a per-pixel visibility test process was implemented using GLSL on two devices:

1. A desktop PC with an Intel Core i7-3930K (3.20GHz, 6 cores), 56.0GB of RAM, and an NVIDIA GeForce GTX 690 (2 GB of texture memory).

2. A tablet device equipped with an Intel Core i7 3667U (2 GHz, 2 cores), 8 GB of RAM, and an Intel HD Graphics 4000 CPU-integrated graphics processor.

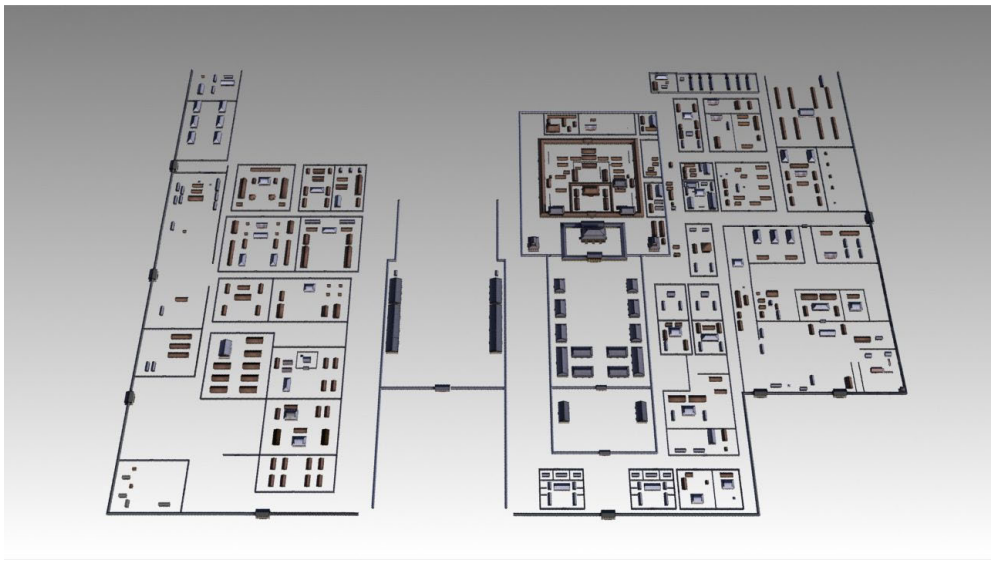


Figure 4.7. 3D models of the virtual palaces in the Heijo-kyo capital. (Courtesy of Toppan Printing Co., Ltd.)

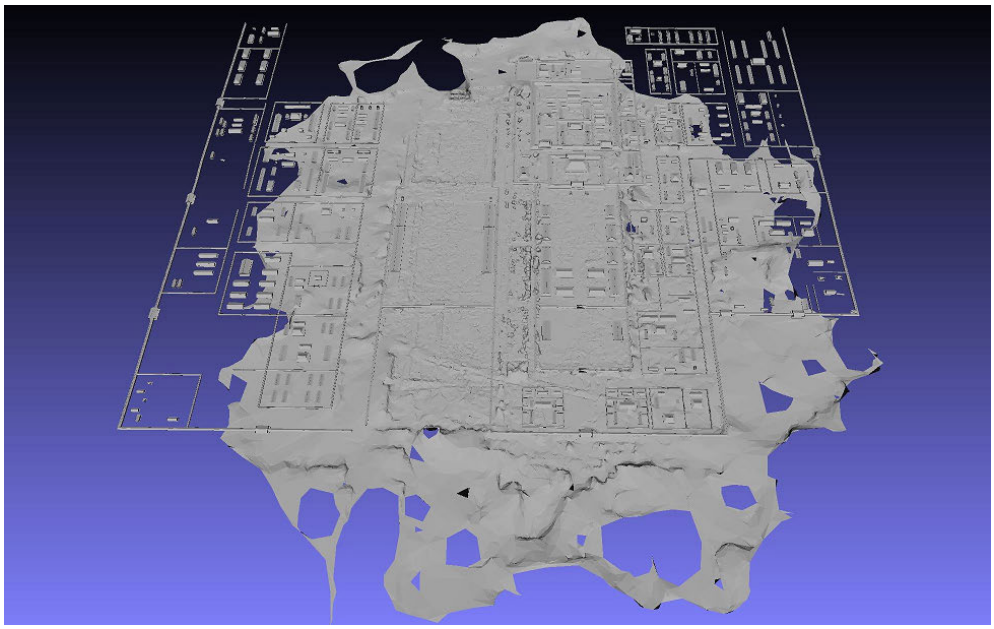


Figure 4.8. Combined and simplified 3D models of the real world and the virtual world, used for the online process.

(a)

(b)

(c)

(d)

Figure 4.9. Free-viewpoint images from various viewpoints, which are shown in Figure 4.6, with and without the photorealistic superimposition of virtual objects.

The online process was carried out on these devices faster than 600 fps and 60 fps, respectively. A notable performance improvement was demonstrated compared with the offline VDTM, which performed at less than 1 fps with a CUDA implementation on the same desktop PC used during the online process.

## 4.5. Discussions

### 4.5.1 Quality of Real-World Images on Structured Viewpoints

We compared the real-world scenes generated from a grid point where the input spherical images were densely captured, and from a grid point generated far from any captured points, because the density of the input spherical images may affect the quality of pre-generated real-world scenes. Figure 4.10 shows examples of spherical real-world images that were pre-generated at the grid points. Figure 4.10(a) shows a generated viewpoint where the input images were densely captured ((0, 0) in Figure 4.6); the viewpoint was successfully generated without artifacts owing to free-viewpoint image generation. Blurring appears in the images in Figure 4.10(b), which is far from any captured points ((-200, -200) in Figure 4.6), because of the errors in the reconstructed 3D surfaces and the low-resolution textures that were projected from the distant cameras. This indicates that planning the capturing positions of the images is also important to improve the appearance of the real-world virtualization in our framework. For 3D reconstruction using automated vehicles, approaches to determine the next-best view and plan efficient capturing paths have been studied [Pit99,BL06]. Such approaches can be adapted for data acquisition for free-viewpoint image generation with further investigation on the quality of the generated images.

### 4.5.2 Quality of Virtual-Object Superimposition of Offline Process

We conducted a small subjective evaluation of the quality of the offline superimposition of virtual objects. We compared an augmented video sequence generated

(a) Viewpoint within densely captured points at position (0, 0) in Figure 4.6.

(b) Viewpoint far from captured points at (-200, -200) in Figure 4.6.

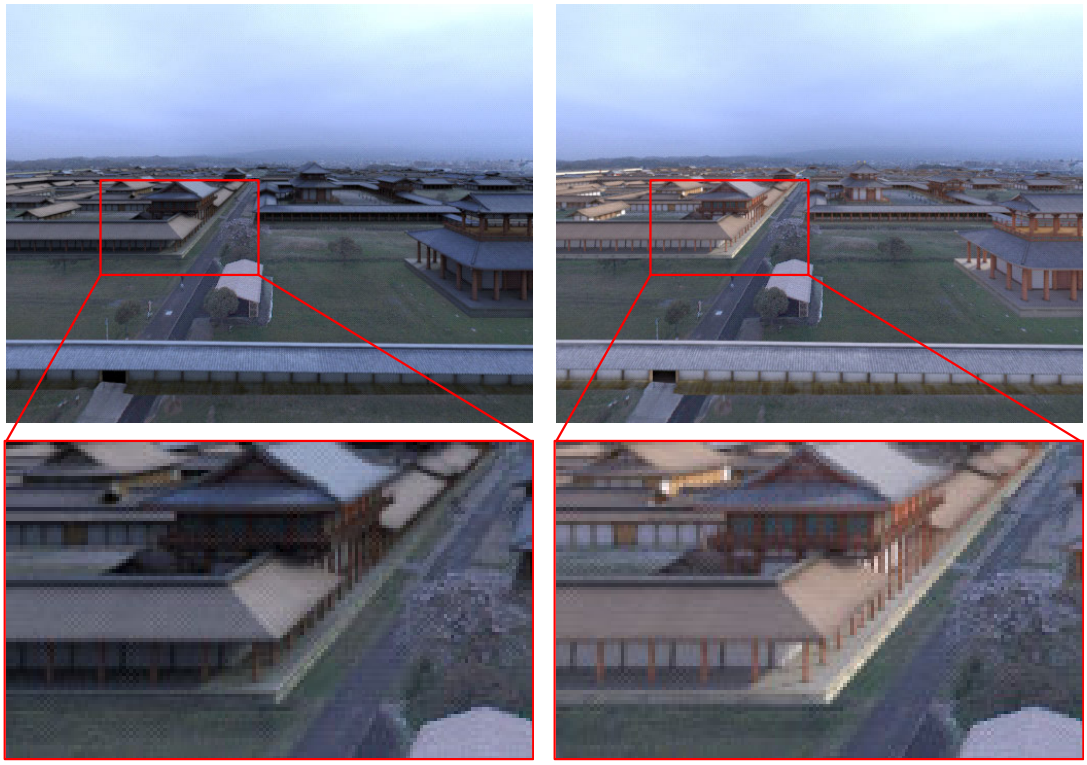
Figure 4.10. Spherical real-world images generated at grid points.

by IBL using an environmental map that was completed by our application and a conventional sequence that was created using a light parallel to the angle of the sun as calculated based on the date and time of the captured sequence. Note that, to eliminate the effects of free-viewpoint image generation and evaluate the superimposition quality by itself, the free-viewpoint image generation technique was not used to generate the sequences used in this evaluation. Figure 4.11 shows a frame of the sequence used in the evaluation. After watching both sequences, ten examinees in their twenties or thirties, evaluated the level of naturalness (1: highly unnatural, 5: highly natural) of the synthesis of real and virtual objects. As a result, the sequence generated using IBL and a completed environmental map received an average score of 4.4, while the conventional sequence scored an average of 2.1., which are significantly different ($p < 0.01$, t-test). The process for offline rendering based on IBL was effective against the use of an ordinary illumination environment, in which the strength and distribution of the skylight were not carefully considered.

### 4.5.3 Density of Structured Viewpoints

Although the use of a larger interval of structured viewpoints can reduce the storage usage during the online process, it can degrade the appearance of the resultant images. Table 4.1 shows the relation between grid sizes and amount of images. Because the proposed rendering requires huge data amount in small grid size, it is important to investigate the trade-off between the data amount and image quality, and find useful trends to determine an appropriate grid size. We performed a subjective experiment to investigate the effects of viewpoint density by changing the grid size in the application described in Section 4.3. Such subjective evaluation should be preferable than objective evaluations, e.g., comparing SNR between images generated using different grid sizes, because we intend to investigate the effects for human vision rather than the objective image quality.

In this experiment, 14 examinees in their twenties or thirties scored the naturalness of video sequences using the interface shown in Figure 4.12. The augmented videos shown to the examinees were generated by our application using six different grid sizes: 10, 20, 40, 60, 80, and 100 m. In addition, we prepared a



(a) Using full spherical image as illumination: Scored an average of 4.4.

(b) Using only a parallel light: Scored an average of 2.1.

Figure 4.11. A sample frame in sequences used for an evaluation of the offline superimposition. Ten examinees evaluated the naturalness of the synthesis of real and virtual objects in five scales (1: highly unnatural, 5: highly natural).

reference video without using the structured viewpoints; i.e., virtual objects were directly rendered onto every frame in the video sequences, in which real-world scenes were generated by VDTM frame-by-frame in offline. It is equivalent to generate infinite number of grid points using 0 m grid size. We prepared two types of view direction for each grid size: downward and horizontal direction. Sampled frames from each video are shown in Figures 4.13 and 4.14. The examinees scored each sequence on a scale of 0 to 100 using a trackbar interface, regarding the score of reference video as 50.

Figure 4.15 shows average scores of the naturalness by the experiment. To remove outliers, 20% of scores (the highest 10 % and the lowest 10 %, respectively) were ignored to calculate the average and the standard deviation, as well as performing a multiple comparison test. The naturalness decreased with larger grid size. We performed a Williams multiple comparison test ($p < 0.05$), in which the control group was set as the score of the reference video, assuming the score distribution is monotonically decreasing. Scores of which grid size is larger than 60 m, were significantly unnatural comparing with those of reference video, which was generated with grid size 0 m. This trend was same between horizontal and downward view direction. Results by grid size 10 or 20 m surprisingly demonstrates better quality than the reference video; it is thought that the flicker appeared on the reference video, because the blending weight among real-world textures determined by VDTM frequently changed. 20 m grid size, which was used for our experiment described in Section 4.4, showed the performance similar to using smaller (i.e., 10 m) grid as well as using the reference video, under our experimental environment. Although this comparison result cannot be directly used for determining the best grid size, which is a trade-off between the quality and the number of pre-generated images, it intends that the larger grid size (i.e., larger than 60 m in this experiment) causes significant negative effects to user experiences. Figure 4.16 indicates the negative effects using a larger grid,

Table 4.1. Amount of images required to prepare a 400m × 400m grid structure. The structured views are assumed to be stored as 8-bit 4-channel (RGB and depth) cube-maps of 1024 × 1024 pixels.

| Grid size | # of grid points | Data amount (not compressed) |
|---|---|---|
| 10m | 1681 | 40.3 GB |
| 20m | 441 | 10.6 GB |
| 40m | 121 | 2.9 GB |
| 80m | 36 | 864 MB |
| 100m | 25 | 600 MB |
| 200m | 9 | 216 MB |

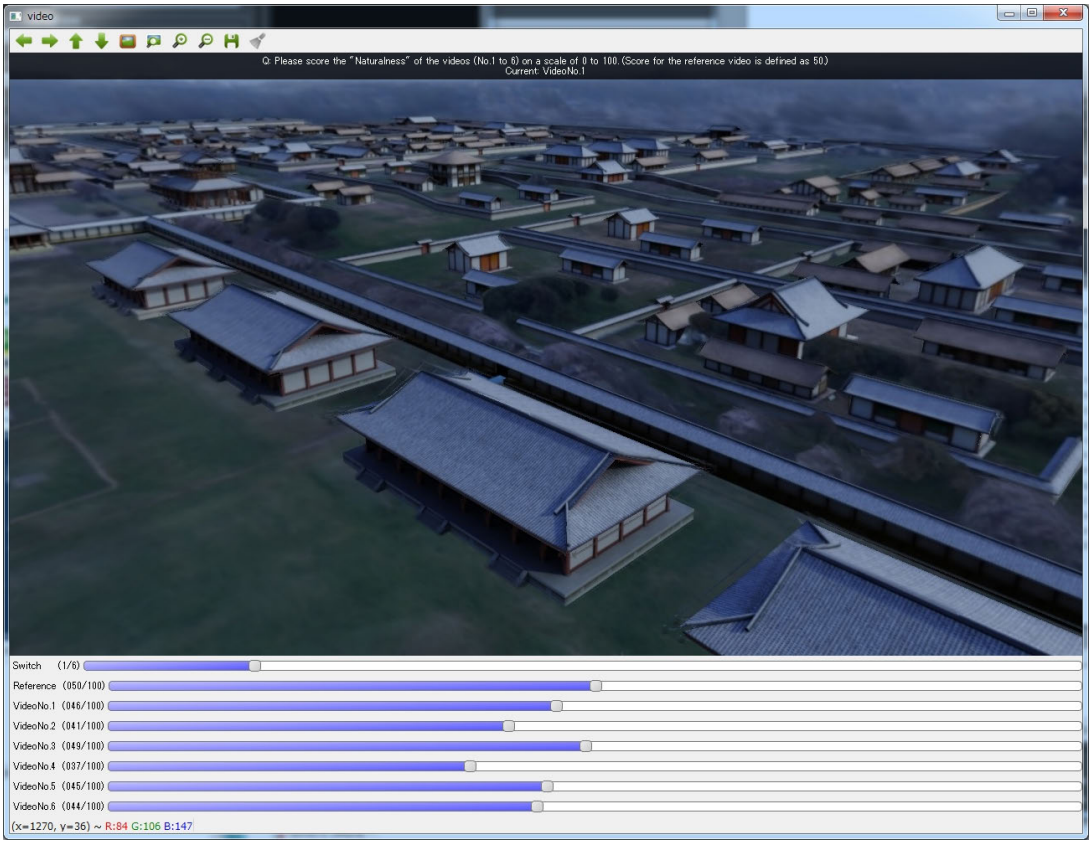



Figure 4.12. Interface used for evaluation using video sequence with a change of grid size. Examinees evaluated the naturalness of the sequence in a scale of 0 to 100.

i.e., the low resolution of the entire scene, and the skewing/blurring of the virtual building posts. When a larger grid was used, the resolution of the resultant images was lower owing to the textures projected from the distant viewpoints. The skewing and blurring posts of the buildings were due to the simplified 3D models, in which some of the vertices were removed, and the effects became more pronounced in the larger grid. In addition, a significantly large grid occasionally generated clearly incorrect textures because some surfaces of the complex 3D shapes were not visible from all four neighboring grid points.

The trend of the relation between the quality and grid sizes basically varies depending on the environment: the complexity of the model and the distance to

(a) Reference video (0 m grid size).

(b) Grid size: 10 m.

(c) Grid size: 20 m.

(d) Grid size: 40 m.

(e) Grid size: 60 m.

(f) Grid size: 80 m.

(g) Grid size: 100 m.

Figure 4.13. Augmented views from the same viewpoint with a change in grid size (horizontal view direction).

(a) Reference video (0 m grid size).

(b) Grid size: 10 m.

(c) Grid size: 20 m.

(d) Grid size: 40 m.

(e) Grid size: 60 m.

(f) Grid size: 80 m.

(g) Grid size: 100 m.

Figure 4.14. Augmented views from the same viewpoint with a change in grid size (downward view direction).

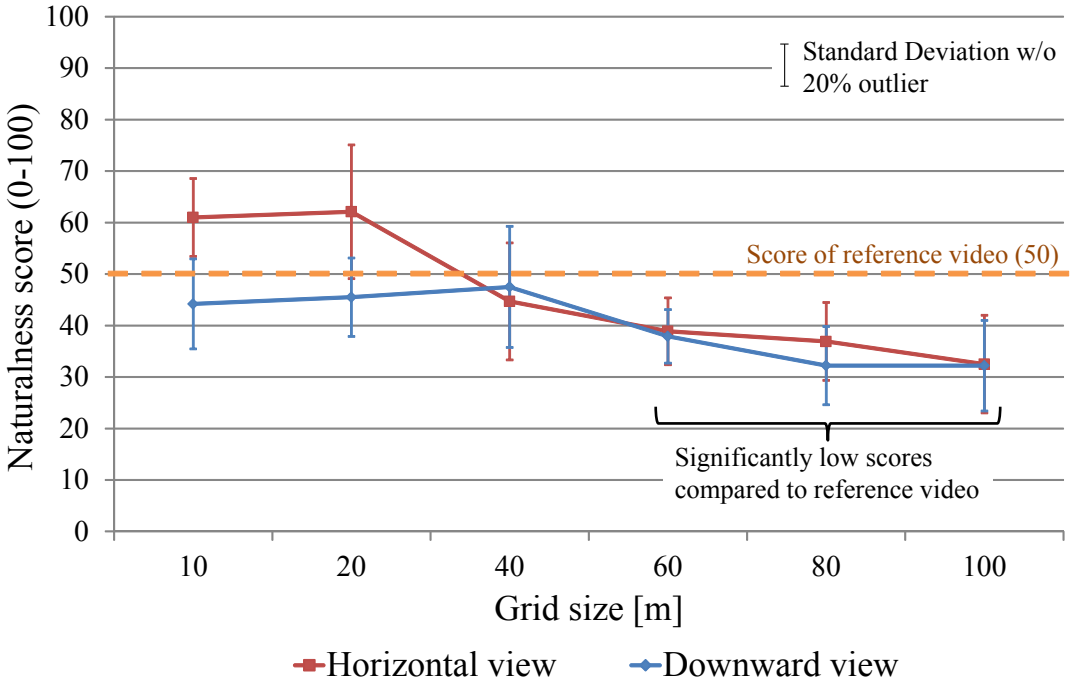



Figure 4.15. Naturalness scores of augmented video sequences with a change in grid size.

the objects. Surfaces on the 3D shapes that are not visible from all neighboring grid points, can be calculated preliminarily; and further, they occupy a large area on the free-viewpoint image when the distance to the surface is small. As a trade-off between a reduction in the number of images and the appearance of the generated images, a promising solution is to change the density of the structured viewpoints, depending on the complexity of the model and the distance to the objects, using quad-trees or similar structures.

### 4.5.4 Limitations

The proposed framework has some limitations. First, the number of pre-generated images can be large because of the offline rendering of virtual objects for multi-viewpoint images. This issue is common to free-viewpoint image generation, particularly for IBR. To address this, multi-viewpoint image compression ap-


Figure 4.16. Negative effects using a larger grid sizes: Closeups from Figures 4.13 and 4.14.

proaches [MRG03, SKC03, LLW04] for IBR can be employed in the proposed framework. Second, although dynamic objects can be superimposed by this framework if the objects are initially rendered, such use is impractical. The third limitation is that this framework do not accommodate real-time illumination change in the illumination environment in a straight-forward manner. To overcome this, certain AR photometric registration methods can be employed, such as image synthesis from the appearances rendered under various illumination environments [SHK+05].

The proposed rendering framework can be potentially employed for MR-world exploration applications using a static environment, such as Google SketchUp. Although there are limitations for use under a dynamic environment, the proposed rendering framework can be a promising approach for a variety of applications,

such as a landscape simulation. The historical application introduced in Section 4.3 is a typical example.

## 4.6. Summary

This chapter described a novel framework for the photorealistic superimposition of static virtual objects into the setting of a real-world virtualization. Offline rendering and a free-viewpoint image generation were combined to preserve the quality of offline rendering in a free-viewpoint MR environment, which provides the user with freely configurable viewpoints. The computational cost of the online process is highly reduced through the pre-generation of structured viewpoints and the pre-rendering of virtual objects at these viewpoints.

We introduced a practical implementation for fly-through application based on the proposed framework, which inputs spherical images captured through the method described in Section 3.2. The application provides a free configuration of the user's viewpoints on a 2D plane through a VDTM-based free-viewpoint generation technique using pre-generated viewpoints at the grid points. In the experiments conducted at a historical site, our implementation demonstrated a high performance. These experiments also illustrated the effects of rendering using IBL in an offline process, which improves the appearance of virtual objects, as well as some problems that should be examined further, such as the determination of the most effective grid sizes. Increasing the dimension easily leads to an enormous number of images. This issue should be considered when developing large applications. In Section 5.5, two applications based on this implementation are described.

# Chapter 5

# Photorealistic MR Applications Based on Proposed Framework

## 5.1. Overview

The proposed framework can accommodate various specific imaging and rendering methods. Using the proposed imaging and rendering frameworks either fully or partly, many kinds of MR applications can be developed as discussed in Section 2.3. This chapter describes following four applications with application-specific discussions and evaluations:

1. HDR immersive panorama
   Full spherical aerial HDR images can be employed in immersive panorama applications by applying perspective projection of sphere-mapped spherical images. However, we should appropriately compress the dynamic range when displaying HDR images for ordinary display devices. We developed a practical immersive panorama application and investigated different display methods using dynamic range compression.

2. Augmented immersive panorama
   By directly superimposing virtual objects into full spherical aerial images without the use of a real-world virtualization, the augmented spherical images can be used for augmented immersive panorama applications. This application does not provide a change in viewpoint other than at the location

where the images were initially captured. A historical entertainment application has been developed based on an augmented immersive panorama for use in a public demonstration.

3. Real-time AR
   With some modifications, the proposed rendering framework can be employed for (video see-through) real-time AR applications of real-world scenes captured in real-time, instead of a real-world virtualization. A simple real-time AR application was developed based on 1D structured viewpoints and an image-based free-viewpoint image generation method.

4. MR-world interactive exploration
   When the proposed framework is fully employed, the produced application realizes the capability of interactive MR-world exploration. An example of the implementation described in Section 4.3, is easily extended to other situations with a small modification. We developed two applications: an extension of the implementation to use 3D cubic structure, and a simple application for the virtual walk-through using 1D structure based on an ordinary spherical imaging using an OMS during walking on the ground.

## 5.2. HDR Immersive Panorama

The full spherical aerial HDR images generated in Section 3.3 can be used for immersive panorama applications, which allow users to freely look around the scene from the location where the images were captured. This section describes an immersive panorama application based on an HMD and a head tracker using the spherical HDR images. To show HDR images on ordinary displays, the radiance has to be compressed through a process referred to as tone-mapping. Numerous tone-mapping methods have been developed [RHD+10] to search for suitable approaches that successfully compress the radiance of an image while saving its specific textures and without generating unnatural artifacts. In most tone-mapping methods, HDR images are converted using parameters determined from the whole image, such as the maximum and logarithmic mean intensities. However, immersive panorama applications generally crop and convert a spheri-

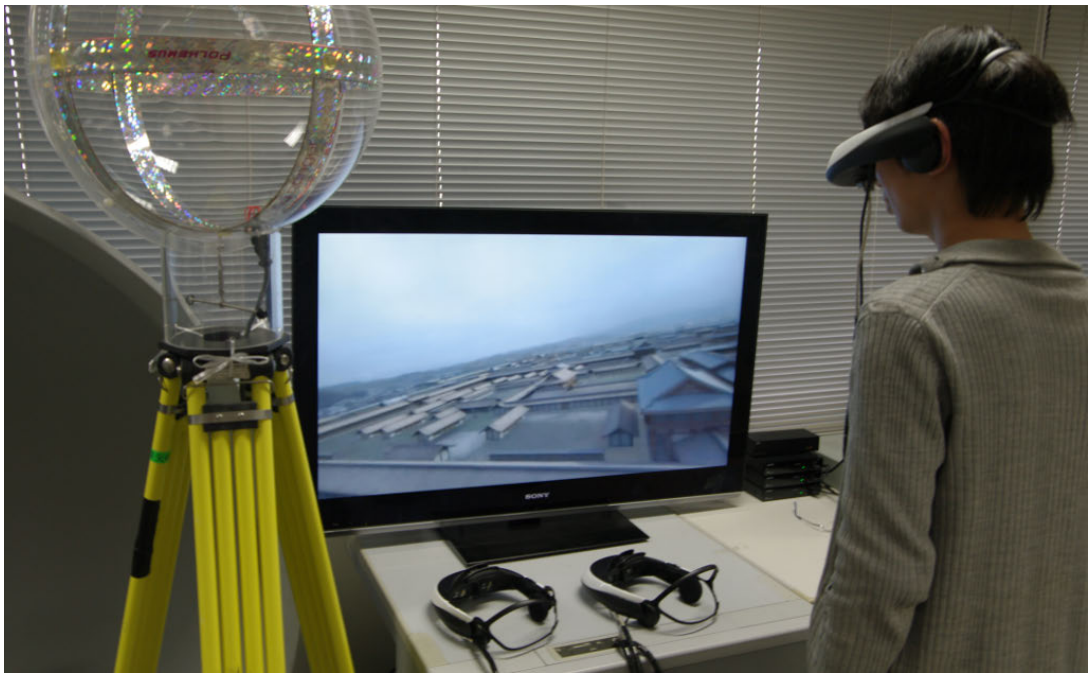
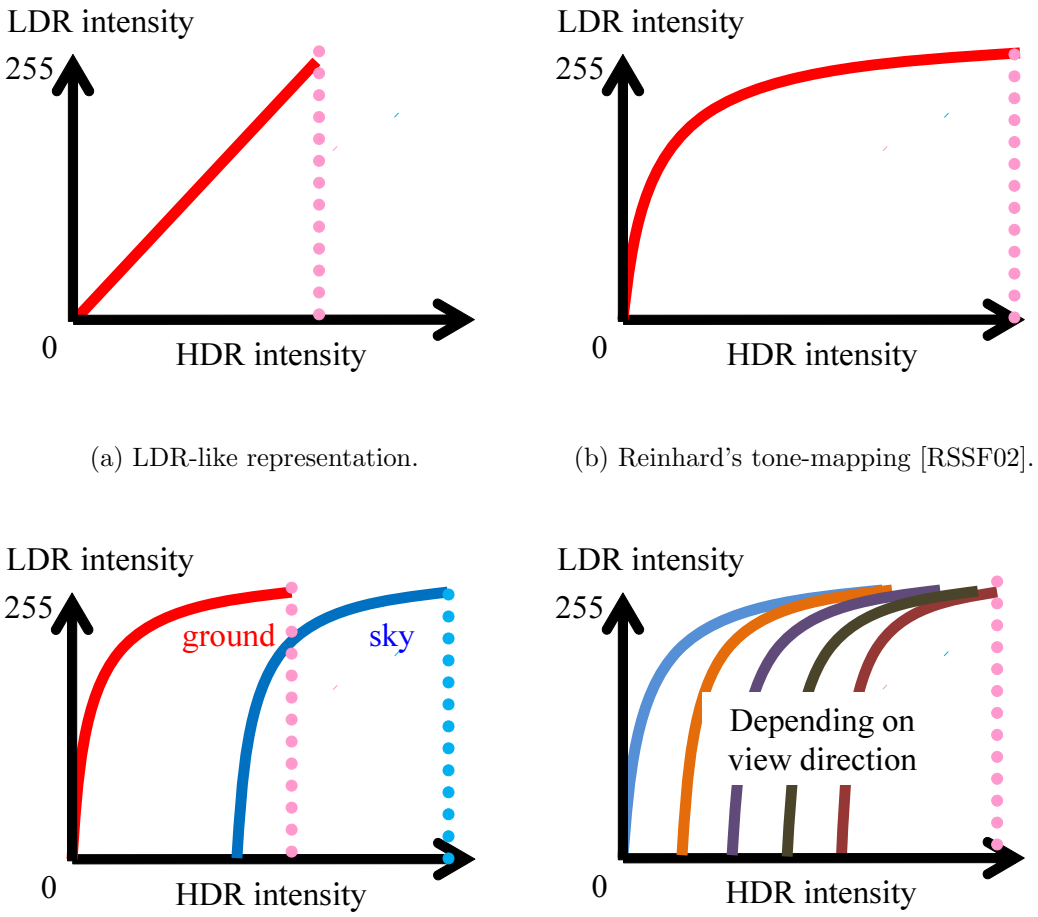


Figure 5.1. Immersive system used in the experiments.

cal image into a perspective image such that the entire image is not displayed all at once. There may be more suitable tone-mapping methods than those used for usual perspective images. It is therefore important to investigate tone-mapping approaches suitable for such applications to improve the fidelity of scenes with large dynamic ranges.

### 5.2.1 Immersive Panorama System with HMD and Head Tracker

We conducted a subjective evaluation of four display methods using an immersive system comprising an HMD and a head tracking device [OYTY98]. Using this system, a spherical image is converted into a perspective image in real-time based on the user's viewing direction, as detected by an electro-magnetic sensor (Fastrak, Polhemus) mounted on the HMD (HMZ-T1, Sony Corp.). Figure 5.1 shows the system used in our experiments. From the location where the spherical image




(a) LDR-like representation.

(b) Reinhard's tone-mapping [RSSF02].

(c) Region-wise tone-mapping.

(d) View direction dependent tone-mapping.

Figure 5.2. Illustrative tone curves for four display methods.

was captured, users can freely change their viewing direction by changing their head pose; i.e., the viewing direction corresponds to the user's head direction in real-time. We used spherical aerial HDR images using taken from OMSs using the method described in Section 3.3 for an immersive panorama system.

The four display methods outlined below were evaluated. Figure 5.2 illustrates an example of the tone curves for each of these methods.

1. LDR-like representation
   This is a luminance transformation method that uses a linear tone curve, such as that shown in Figure 5.2(a), and generates an LDR image with the same appearance as that generated through traditional LDR imaging.

2. Reinhard's tone-mapping [RSSF02]
   This uses the maximum and log-mean intensities to tone-map the HDR image. It is one of the most popular tone-mapping methods.

3. Region-wise tone-mapping
   Spherical aerial images are divided into bright sky and dark ground regions. Using the method of [KOKY12], these regions are then tone-mapped independently using different parameters.

4. View-direction dependent tone-mapping
   An input HDR image is tone-mapped based on the intensity of the field of view of a perspective image converted using an immersive system based on the user's viewing direction. The intensity of a position in the scene varies according to this viewing direction.

### 5.2.2 Display Methods for Spherical Aerial HDR Images

Let us now look at the implementation of each display method and the results of a subjective review.

#### LDR-like representation

A linear intensity transformation is used to generate an image that looks the same as any image generated through traditional LDR imaging. A tone curve can be defined as

$$I_l = \begin{cases} 1.0 & (I_h > 1.0) \\ 2^v I_h & (otherwise), \end{cases} \tag{5.1}$$

where $I_h$ is the intensity in an HDR image, $I_l$ denotes the intensity of an LDR image normalized to the range $[0, 1]$, and $v$ is the exposure value of the generated LDR image. We determined $v$ to reduce overexposed and underexposed pixels

Figure 5.3. Image generated through LDR-like representation.

in the generated LDR image. For this experiment, underexposure is defined as $I_l < \frac{1}{16}$, and overexposure as $I_l = 1.0$. Note that, as shown in Figure 5.3, many overexposed and underexposed pixels remain in the resulting image, which was converted from an HDR image captured as described in Section 3.3.

### Reinhard's tone-mapping [RSSF02]

The tone-mapping method proposed by Reinhard et al. [RSSF02] is composed of two steps. In the first step, the intensity $I_h$ of an HDR image is transformed into LDR intensity $I_l$ by a function involving the log-mean $\bar{I_h}$:

$$I_s = \frac{\alpha}{\bar{I_h}} I_h, \tag{5.2}$$

$$I_l = \frac{I_s \left(1 + \frac{I_s}{I_{max}^2}\right)}{1 + I_s}. \tag{5.3}$$

Note that $I_s$ is a scaled intensity using scale factor $\alpha$, and $I_{max}$ denotes the maximum intensity scaled using Equation (5.2). The second step, so-called "dodging-and-burning" is applied to areas of low and high intensity to improve their visibility (detailed in [RSSF02]). Parameter y is automatically estimated by [Rei02]. Figure 5.4 shows the result of tone-mapping the same frame used in Figure 5.3.

Figure 5.4. Result using Reinhard's tone-mapping method [RSSF02].

### Region-wise tone-mapping

Dividing images into sky and ground regions, we applied the approach proposed by Kitaura et al. [KOKY12] to tone-map these regions using separate parameters. To fully automate this process, we improved it for spherical aerial images. Specifically, we used *a priori* knowledge of aerial images to support automatic GrabCut segmentation [RKB04] in place of a user input of the seed pixels. We assumed that the camera pose was estimated by SfM in the same fashion as described in Section 3.3.5, and that the captured HDR image was aligned to fit the horizon in the scene to the horizontal scan line of the image. Under these assumptions, the seed pixels for the dark (ground) region can be taken from the lower-half of the image, as shown in Figure 5.5. The distribution of seed pixels for the bright (sky) region depends on the landscape of the scene. We used the upper-third of the image as seed pixels for the bright region, resulting in a successful separation of the sky from the ground, as shown in Figure 5.6.

Region-wise tone-mapped images, such as that shown in Figure 5.7, are then produced by applying Reinhard's tone-mapping [RSSF02] to each region independently.

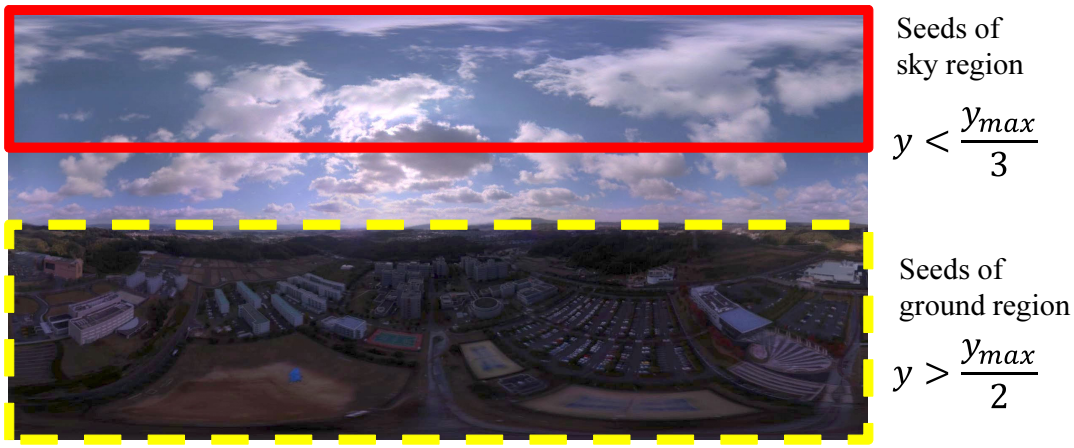



Figure 5.5. Seed pixels for GrabCut [RKB04]: $y$ denotes the vertical component of the image coordinates; and $y_{max}$ is the height of the image.

Figure 5.6. Regions determined using GrabCut [RKB04].

Figure 5.7. Result of region-wise tone-mapping.

### View-direction dependent tone-mapping

For immersive panorama applications in particular, it is worth investigating tone-mapping methods whose tone curves vary according to the user's viewing direction. We applied a new approach similar to the tone-mapping for HDR video [KUWS03]. This approach consists of two processes:

1. Offline process: Pre-calculation of log-mean and maximum intensities.
2. Online process: Tone-mapping using these pre-calculated parameters.

The log-mean and maximum intensities were calculated from planer perspective images generated for each viewing direction (in this study, we calculated the parameters for every one degree of latitude and longitude). Equations (5.2) and (5.3) were then applied for tone-mapping in real-time using the calculated parameters and $\alpha = 0.18$, which is the default value used in Reinhard's tone-mapping [RSSF02]. Note that although this experiment used an offline process, it is possible to implement a fully real-time process by leveraging the GPU computation, as in other real-time tone-mapping methods [GWWH03]. The results, shown in Figure 5.8, reveal how the intensity at a given position in a scene varies according to the user's viewing direction.


Figure 5.8. Examples of view-direction dependent tone-mapping: The intensity of the same position in the scene varies according to the viewing direction.

### 5.2.3 Subjective Evaluation Using Immersive System

To investigate the features of the four display methods described above for spherical aerial HDR images, we conducted a subjective evaluation using an immersive system. Examinees used an HMD to view scenes generated using these four methods, and responded to each of the following questions with a rating of one (worst) to five (best):

**Q1:** Were the scene and textures easily recognizable? (Visibility)

**Q2:** Did the tone, brightness, and appearance of the textures appear to be natural? (Naturalness)

**Q3:** Did you get the feeling that you were actually flying through the sky? (Immersion)

**Q4:** Were you satisfied with the display method as a whole? (Total)

Note that these questions were originally given in Japanese. The examinees were ten people in their twenties or thirties; some of them were familiar with the VR and image processing techniques. The examinees were also allowed to comment freely after answering the questionnaire. The order in which the four display methods were presented was randomized, and the examinees were allowed to switch the display mode among the representations as many times as they wished.

In this experiment, we stated following two hypotheses:

1. LDR-like tone-mapping, which does not represent high dynamic range intensity, is evaluated worth than other methods. The representable dynamic range directly affects the visibility of the scene (Q1); while the degree of satisfaction (Q4) should demonstrate the similar trends. It intends that the combination of HDR imaging and appropriate representation are effective for immersive panorama applications.

2. Region-wise and view-direction dependent tone-mapping are scored higher than to tone-map a whole spherical image by Reinhard's tone-mapping, because of its advantages in the visibility.

The results of the evaluation are shown in Figure 5.9. This figure also indicates the pairs that had a significant difference ($p < 0.05$ and $p < 0.01$), as calculated using the Tukey's test. In total, Reinhard's method and the view-direction dependent method had higher ratings than the LDR-like method. Region-wise tone-mapping was significantly lower than the other three, especially with regard to naturalness.

### 5.2.4 Discussions

#### Effects of HDR imaging for immersive panorama

The results demonstrate that the first hypothesis is correct with some exceptions. For Q4 in Section 5.2.3, with the exception of region-wise tone-mapping, nonlinear tone-mapping methods had significantly higher ratings than LDR-like representation methods (as shown in Figure 5.9). We can therefore confirm that HDR imaging and appropriate display methods are superior for immersive panoramic applications, as well as for traditional uses of HDR images such as conventional photography and IBL. This superiority is most noticeable in the reduction of overexposed and underexposed pixels. Furthermore, since immersion is one of the chief goals of virtual reality [Moe97], it is noteworthy that HDR imaging does not decrease the immersion despite the use of a nonlinear intensity transformation in the tone-mapping, which can be regarded as unnatural.

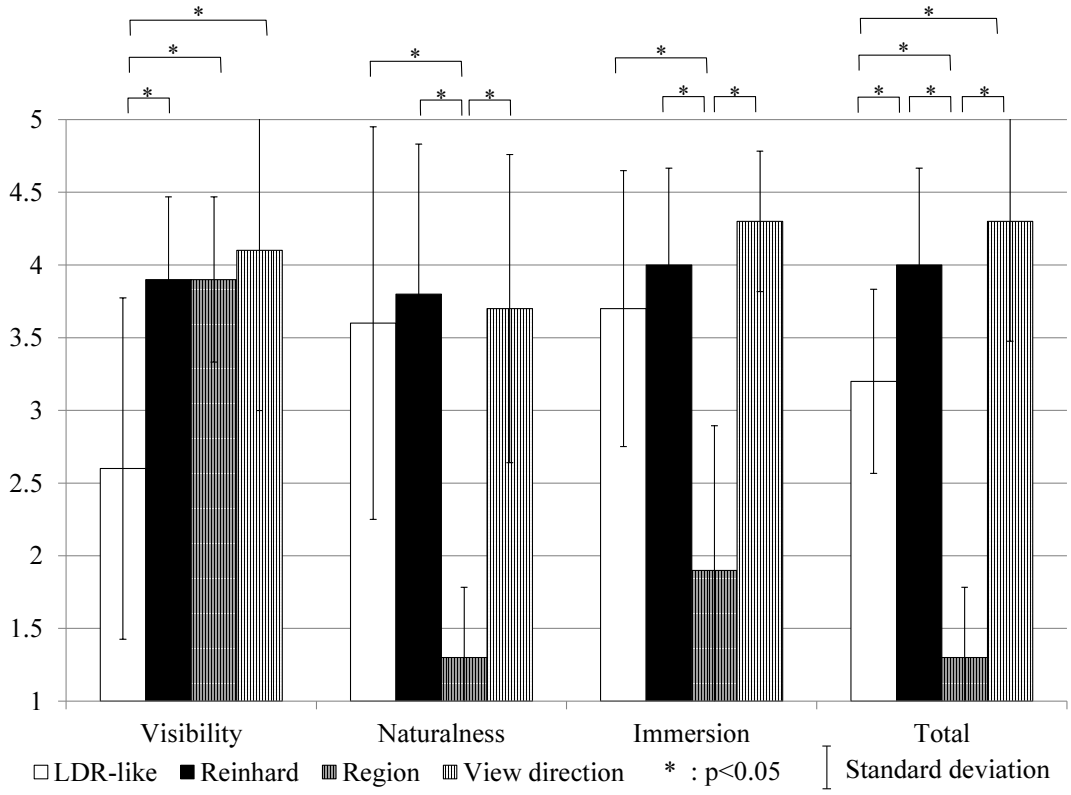



Figure 5.9. Results of subjective experiment involving the four display methods.

### The best HDR image display method for immersive panorama

This experiment partially refuted our second hypothesis. Although region-wise tone-mapping had a higher visibility rating (along with other HDR tone-mapping methods), significantly lower ratings were obtained for the other questions. Free comments from the examinees suggested that the most critical factor in determining the naturalness was the darkness of the sky relative to the intensity of the ground.

Contrary to the expectation, there are no significant differences between Reinhard's tone-mapping method and the view-direction dependent tone-mapping. Free commentary from the examinees indicated that the view-direction dependent tone-mapping produced unnatural intensity variations as the viewing direction changed in real-time. This trend would change under different applications

and target environments; e.g., if users stay watching groundward scenes with little change of the luminance, the artifacts by view-direction dependent one does not occur. It indicates that the parameters related to human factors, such as acceptable rates of intensity variation, should be investigated to guide further development. If it can overcome this perceived deficiency, or can find suitable applications, view-direction dependent tone-mapping may become a more popular display method.

## 5.3. Augmented Immersive Panorama

By directly rendering virtual objects into spherical images, an augmented immersive panorama can be developed. Such rendering is considered for applications using the proposed imaging framework, as well as the offline rendering process in the proposed rendering framework without an explicit real-world virtualization. To display the augmented spherical images, we can basically employ the same hardware configuration as in an immersive panorama without the superimposition of virtual objects, which provides a change in the user's viewing direction.

### 5.3.1 Augmented Immersive Panorama System Using Full Spherical Aerial Image Sequence

We developed a virtual historical tourism application based on the augmented immersive panorama technique using the full spherical aerial image sequence described in Section 3.2.6. Unlike the free-viewpoint rendering described in Chapter 4, this application does not provide a free configuration of the user's viewpoint, but enables the users to move along the captured path of the spherical image sequence. The virtual objects of the old palaces, which are from the same 3D model as that used in Section 4.3, were rendered using IBL without generating structured viewpoints. The resultant augmented spherical image sequence consists of 1,900 frames, whose real-world scenes are completed by a method described in Section 3.2, and Figure 5.10 shows the frames sampled from this sequence. Note that the selected frames shown in Figure 5.10 correspond to those shown in Section 3.2.6.

(a) First frame.

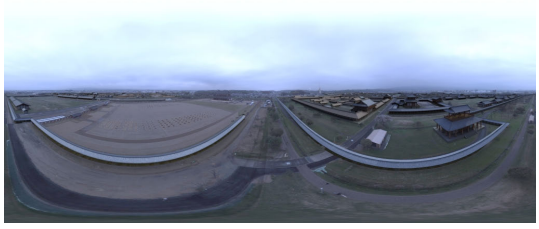
(b) 200th frame.

(c) 400th frame.

(d) 600th frame.

(e) 800th frame.

(f) 1,000th frame.



(g) 1,200th frame.

(h) 1,400th frame.

Figure 5.10. Augmented spherical image sequence for augmented immersive panorama application.

(a) TV monitor system. (b) HMD system.

Figure 5.11. Appearance of augmented immersive panorama systems.

### 5.3.2 Public Experiment

In 2010, an experimental public demonstration using the augmented immersive panorama application was conducted during the commemorative event for the 1,300th anniversary of Nara Heijo-kyo capital for a two-week period, and over one-thousand people experienced the application. We developed two display systems using an HMD and a TV monitor, which are shown in Figure 5.11. The configuration of the HMD system is similar to that used in the immersive panorama system described in Section 5.2, which changes the viewing direction of the augmented scene using the user's head pose measured using a gyroscope mounted on the HMD. In the system using a TV monitor, the viewing direction could be changed using a joystick. The specific configuration of the demonstration system is shown in Table 5.1. Note that in the presented system, the user can switch sequences with and without virtual objects to achieve a better user experience.

We obtained a large amount of feedback from the users during the demonstration, and most of them were positive comments. We also heard many suggestions regarding the demand for annotations of the virtual buildings and animating of virtual objects. From these user suggestions, in addition to realizing a photo-realistic superimposition, a high-quality user interface is also required for these virtual historical tourism systems.

## 5.4. Real-Time AR

We developed a simple real-time AR application using mobile device based on the proposed rendering framework. In our implementation, a user explores AR scenes from viewpoints on initially designated linear paths. Based on the flow illustrated in Figure 5.12, the appearances of virtual objects and depth maps at structured viewpoints were generated offline. During the online process, we employed a morphing technique [MD96], which is an appearance-based free-viewpoint image generation method.

### 5.4.1 Offline Process: Rendering of Virtual Objects at Structured Viewpoints

During the offline process, structured viewpoints are designated on a line every couple of meters. At each viewpoint, the appearances of the virtual objects and depth maps are rendered, as shown in Figure 5.13. A spherical image is acquired from a location near the structured viewpoints; the image is employed as the illumination for the virtual objects with manual adjustments (e.g., manually designating additional light sources).

Table 5.1. System configuration for augmented immersive panorama.

| | Hardware | Model | Specifications |
|---|---|---|---|
| HMD system | PC | DELL Vostro 1400 | CPU: Intel Core2 Duo T7300<br>Graphic: Intel HD Graphics |
| | HMD | iWear VR920 | Resolution: 640×480<br>Vertical FOV: 32°<br>3 DoF gyroscope |
| TV monitor system | PC | Faith MTX2 i7720 | CPU: Intel Core i7 X920<br>Graphic: GeForce GTX 285M |
| | TV Monitor | Sharp AQUOS LC-65RX1W | Display size: 65 inches<br>Resolution: 1920×1080 |

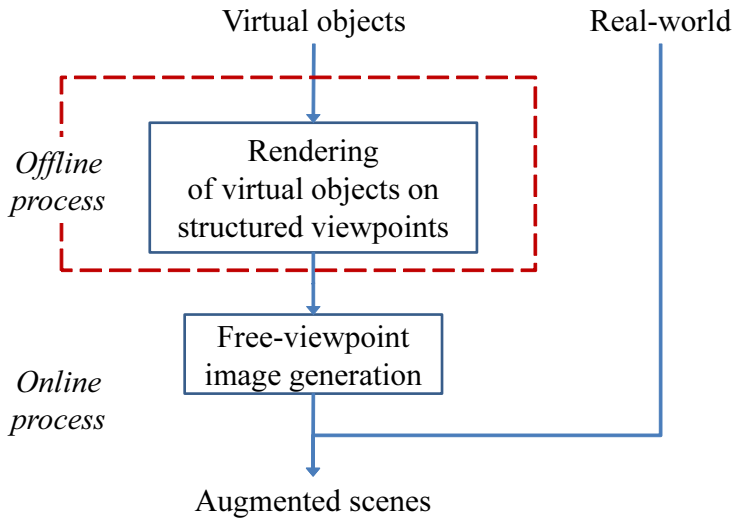



Figure 5.12. Modified rendering framework for real-time AR.

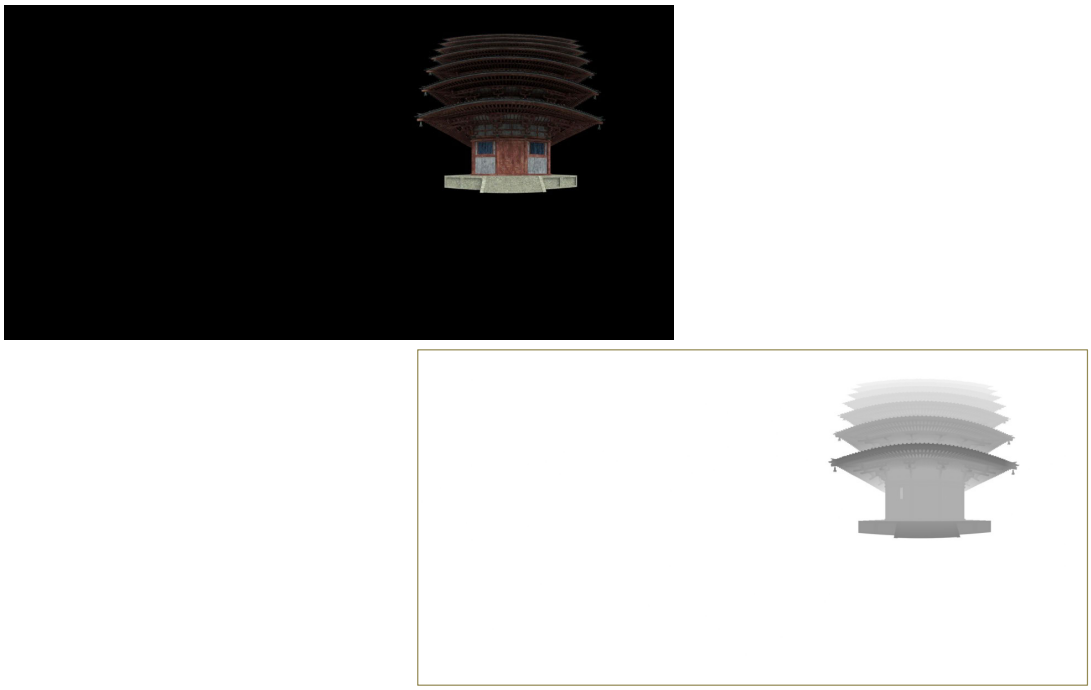


Figure 5.13. Panoramic appearance and depth map of virtual objects at a structured viewpoint.

### 5.4.2 Online Process: Free-Viewpoint Image Generation

The online process superimposes a virtual object into a real scene by morphing [MD96] the two structured viewpoints nearest to the position of the camera capturing the scene. Pre-generated depth maps are used for the 3D shape of the scene, and the appearance of the virtual object is morphed with respect to each pixel of the depth map. This process does not depend on the number of polygons, but rather depends on the number of pixels in the depth map. The morphed textures are blended using a simple weight, i.e., the reciprocal ratio of the distances to the two neighboring viewpoints. The structured viewpoints enable an easy selection of the neighboring views and O(1) calculation of the blending weight. This process is similar to other appearance-based free-viewpoint image generation methods simplified for the use of only two views, such as unstructured lumigraph rendering [BBM+01].

### 5.4.3 Mobile Device Experiment

To investigate the quality of generated scenes and the computational costs of the online process, a basic AR application for a historic site was developed. The application superimposed into a real scene a virtual seven-stair tower, which no longer exists, of the famous Todaiji Temple in Japan. This model was created with reference to the miniature model of Todaiji at the time of its foundation, and is exhibited in the Hall of the Great Buddha, with architectural validation by Yumiko Fukuda, Hiroshima Institute of Technology, and patina expression technology validation by Takeaki Nakajima, Hiroshima City University. The structured viewpoints and captured path of the real scene are shown in Figure 5.14. Eleven structured viewpoints were designated at every 1m, and the virtual object was rendered at every viewpoint using offline GI-rendering software. The virtual object appearance and depth maps were generated as cube maps, for which a total of 66 MB of storage was required. In this experiment, to eliminate the effects of errors in the geometric registration of virtual objects during the online process, the real scenes were preliminarily captured using an OMS, and the captured position and orientation were preliminarily estimated using SfM [Wu13]. In this experiment, a perspective-image sequence generated from a spherical video was

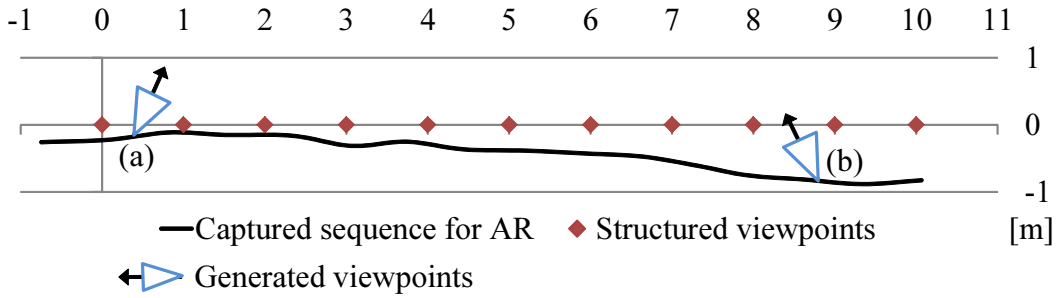



Figure 5.14. Top view of structured viewpoints where virtual objects are rendered offline, capturing path of real-world spherical images for AR application, and online-generated viewpoints.

used as the input for the online process. Figure 5.15 shows examples of augmented scenes, as well as the real images used as the background scene. The online image generation process was performed at 30 fps with a GLSL implementation on a tablet device equipped with an Intel Core i7 3667U (2 GHz, 2 cores), 8 GB of RAM, and an Intel HD Graphics 4000 CPU-integrated graphics processor.

### 5.4.4 Discussions

#### Image quality from online process

To confirm the quality of the images generated using the online process, we compared the AR scenes with an offline-rendered scene. Figure 5.16(a) shows an online-augmented scene, and Figure 5.16(b) shows the scene generated through offline rendering using the same viewpoint and illumination used in the scene shown in Figure 5.16(a). The minor difference between the two images, as shown in Figure 5.16(c), indicates that our online process successfully generates a similar appearance as the offline-rendered images. Because offline rendering includes special boundary processes, such as anti-aliasing, errors mainly occur in the boundaries between the real and virtual scenes. These processes can be employed for the proposed framework, along with some AR studies involving the registration of image quality between real and virtual objects [FBS06].

(a)

(b)

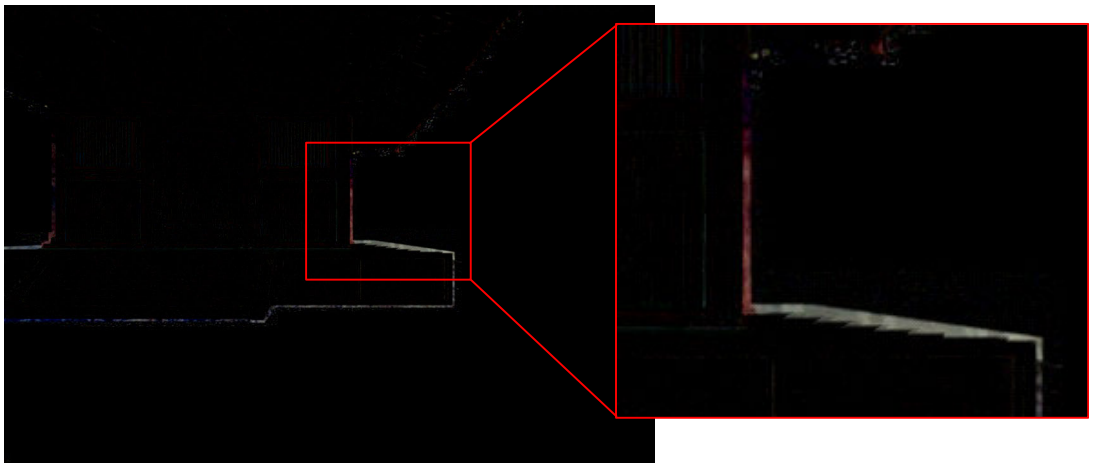
Figure 5.15. Real-time AR based on proposed rendering framework. Scenes with and without superimposition.

### Computational cost

We compared the performance of the proposed framework-based AR application with that of an ordinary rendering method. Our AR application superimposes a virtual model of 1,131,736 polygons, while our implementation does not depend at all on the number of polygons. Table 5.2 describes the performance of the proposed application and the rendering of a textured virtual model, where the latter method does not employ any special lighting effects. The devices used for the performance measurement were the same as those used in the experiment described in Section 4.4. Our application performance was 10- to 100-times faster than the simple rendering of the textured virtual model.

(a) Generated by our online process.

(b) Produced through offline rendering.



(c) Difference between scenes in (a) and (b).

Figure 5.16. Comparison of scenes generated through online free-viewpoint image generation and offline rendering. The online process generates a similar appearance as the offline-rendered image. Errors mainly occur in the boundaries between the real and virtual scenes.

## 5.5. MR-World Interactive Exploration

### 5.5.1 Fly-Through Application using 3D Structure

The application described in Section 4.3 can be extended for using 3D structure by straightforward approach, generating multiple grids in multiple altitude. It should be noted that, this approach requires a larger amount of images than using 2D structure.

#### Implementation

We developed a fly-through application using 3D structure in the same target environment as described in Section 4.3. Augmented free-viewpoint scenes were generated on three grids in multiple altitude, each of these covers 400m $\times$ 400m. In addition to the grid used in the 2D application, whose altitude was set in approximately 50 m from the ground, we also generated grids in 10 and 30 m altitude by the same manner as the offline process of 2D application. From the experiment in Section 4.5.3, there are not a significant degradation of the free-viewpoint images in 40 m grid, the grid points were designated for every 40m $\times$ 40m area, which is larger than that in the implementation in Section 4.3. The horizontal and vertical intervals of structured viewpoints are different (40 m and 20 m). It is because the target environment includes some two-storied buildings; it clearly occurs the missing textures for a large vertical interval of the viewpoints. Note that differences of interval do not increase the complexity

Table 5.2. Frame rates (fps) of the proposed rendering in our AR application and simple rendering of the textured virtual model.

| | Proposed | Textured Model |
|---|---|---|
| Desktop PC w/ NVIDIA GeForce GTX 690 | 480 | 5.1 |
| Tablet PC w/ Intel HD Graphics | 30 | 2.0 |

of the rendering process. In the application using 3D structure, online VDTM blends the textures at eight neighboring structured viewpoints, although the 2D application uses four textures. The eight neighboring structured viewpoints are placed at each vertex of the cube including the user's viewpoint, and they realize O(1) calculations for the blending weight.

#### Results and discussions

Examples of augmented images, which were generated during the online process, are shown in Figure 5.17. This application was implemented in the same desktop PC used in the experiment described in Section 4.4, and achieved similar performance as the 2D application (faster than 600 fps).

Augmented images in low altitude include a lot of visual artifacts. It is thought to be the cause of two problems: 1) The user's viewpoint is relatively close in low altitude. The errors of the shape due to 3D reconstruction and polygon reduction largely affect the appearance of the image, and 2) some parts of complex 3D shapes were not visible from all neighboring structured viewpoints. Artifacts due to the problem 1) are shown in the appearance of occlusion boundary as shown in Figure 5.18. The errors of shape cause blur on the occlusion boundary from scenes viewed from high altitude. The artifacts become larger in low altitude; i.e., clearly incorrect textures were generated. These negative visual effects can be reduced by designating narrow grid; however, it leads to use of an enormous number of images.

### 5.5.2 Walk-Through Application Using Spherical Images Captured on Ground, 1D Structure, and Morphing

This section describes a walk-through application based on the proposed rendering approach. The application is summarized as follows:

- **Intended use**: An MR walk-through application superimposing 3D models of buildings.
- **Input**: Spherical LDR images captured from the ground during moving.

(a) Horizontal location as in Figure 4.9(a).

(b) Horizontal location as in Figure 4.9(b).

(c) Horizontal location as in Figure 4.9(c).

(d) Horizontal location as in Figure 4.9(d).

Figure 5.17. Free-viewpoint images from various altitudes. The altitude is approximately 40 m (left figure), 20 m (right figure) from the ground.

(a) A complex shape viewed from high altitude (approximately 50 m).

(b) A complex shape viewed from low altitude (approximately 15 m).

Figure 5.18. Artifacts on occluding boundary between virtual objects and virtualized real objects.

- **Dimension of structured viewpoint**: One dimensional structure (on a line.)
- **Free-viewpoint image generation**: Morphing technique [MD96] using depth maps from structured viewpoints. Because the target scene is expected to be close to the user's viewpoint, as discussed in Section 2.2.1, appearance-based hybrid rendering approaches should be appropriate for our application. Morphing is a typical appearance-based rendering approach.

In our application, real-world views are first generated at equally-spaced points on a line using a morphing technique. Virtual objects are then photorealistically rendered for every structured viewpoint. It should be noted that, the virtual

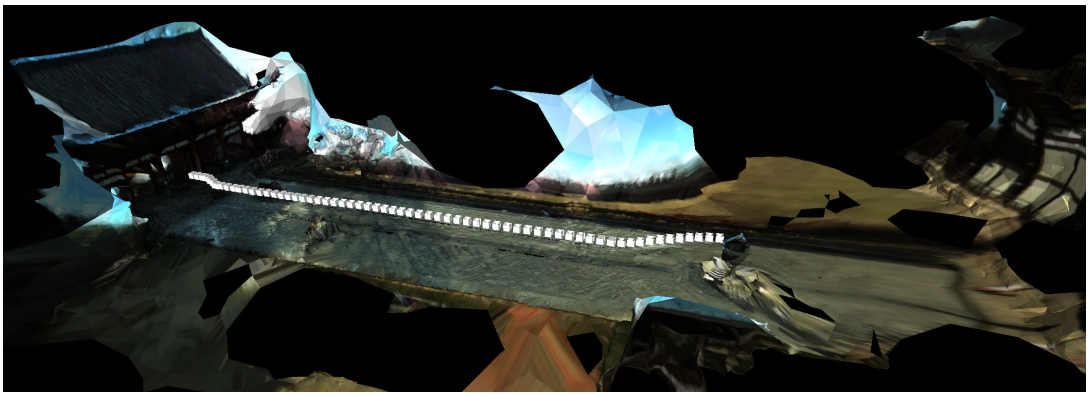


Figure 5.19. Reconstructed real-world 3D models and camera pose of the OMS used.

models and the real environment are not semantically related; there are no any historical meanings of the contents.

### Implementation

Camera pose and real-world dense 3D models are reconstructed from input spherical images, which are captured by an OMS from the ground during walking, by VisualSFM [Wu13] and CMPMVS [JP11], as shown in Figure 5.19. Depth map from each camera is generated from the reconstructed models. Real-scenes from the structured viewpoints are generated by per-pixel morphing [MD96] using the depth map. The images are morphed with respect to each pixel of the depth map. In our implementation, 56 structured viewpoints were designated on a line at every 1m as shown in Figure 5.20. At each viewpoint, the process renders the virtual objects as same manner as described in Section 4.3 (see Figure 5.21), as well as the depth maps from the structured viewpoints. Note that, the application inputs LDR images; additional light sources were manually designated for the rendering.

The online process generates the user's view from the images at the two nearest structured viewpoints from the user's viewpoint by the morphing, as similar fashion as the offline process. This process does not depend on the number of polygons, but rather depends on the number of pixels on the depth map. The

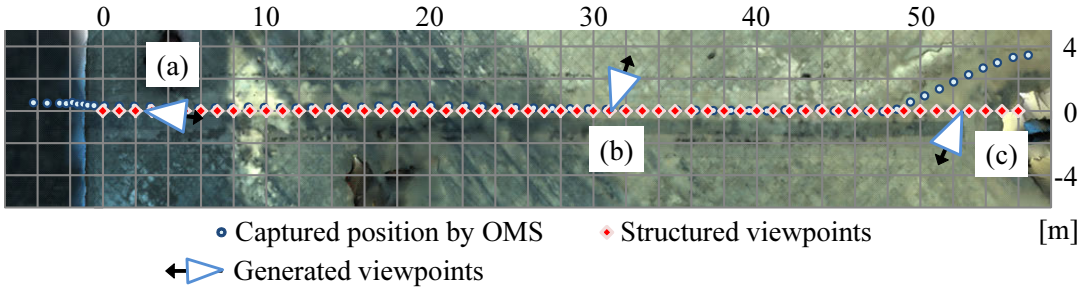



Figure 5.20. Top view of structured viewpoints where MR-world is rendered, capturing path of real-world spherical images, and online-generated viewpoints.

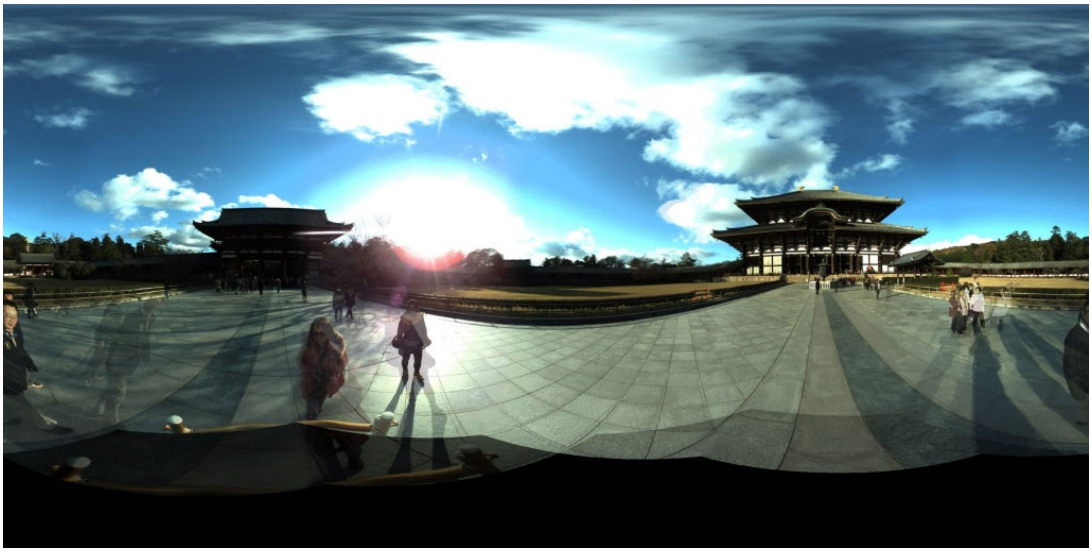
morphed textures are blended using a simple weight, i.e., the reciprocal ratio of the distances to the two neighboring viewpoints. The structured viewpoints enable an easy selection of the neighboring views and O(1) calculation of the blending weight.

### Results and discussions

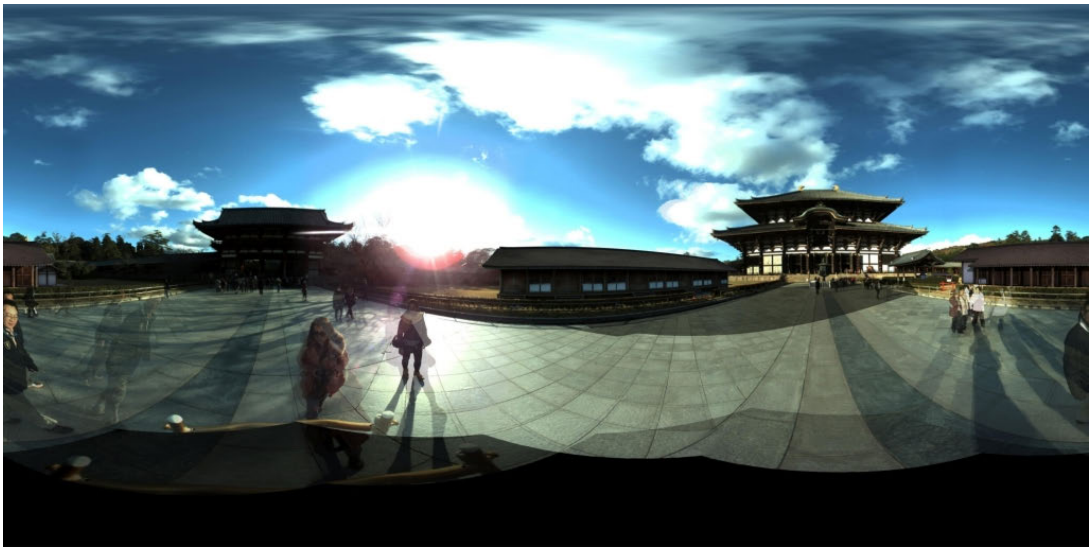
Examples of augmented images, which were generated during the online process, are shown in Figure 5.22. The augmented images were successfully displayed, including occlusion of virtual objects from real hedges. This online process was performed at 30 fps with a GLSL implementation on the same desktop PC used in the experiment described in Section 4.4.

As shown in Figure 5.22(a), the real-world scenes in this application includes the appearance of pedestrians, which can be a cause of reducing the immersive values. There are some methods to remove pedestrian figures from spherical images captured in a fixed point (e.g., [AHK+10]) as discussed in Section 2.1.3. However, it is still a challenging problem for spherical image sequences captured during motion, where this type of image sequences was used in this application as well as Google Street View. A study by Flores et al. [FB10] tackles this problem using Google Street View images captured in environments consisting of planar objects; it can be a future direction to generalize the target scenes of such approaches to arbitrary geometry.

If this application is used around the actual location of the target environ-

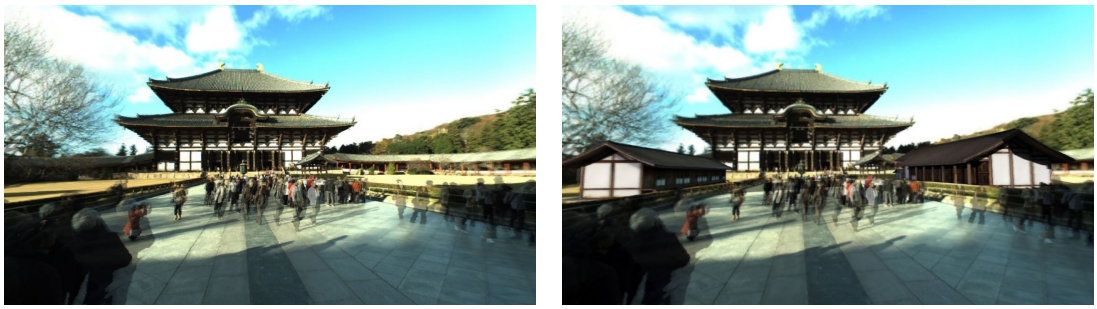


(a) Real-world view at a structured viewpoint.

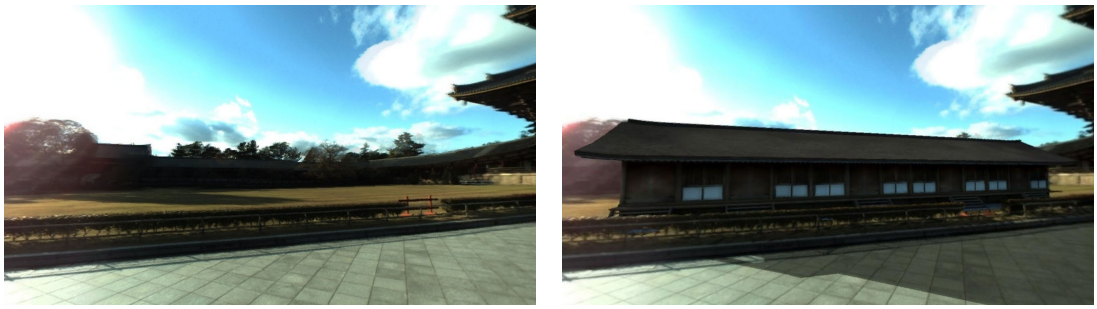


(b) MR-world view at a structured viewpoint.

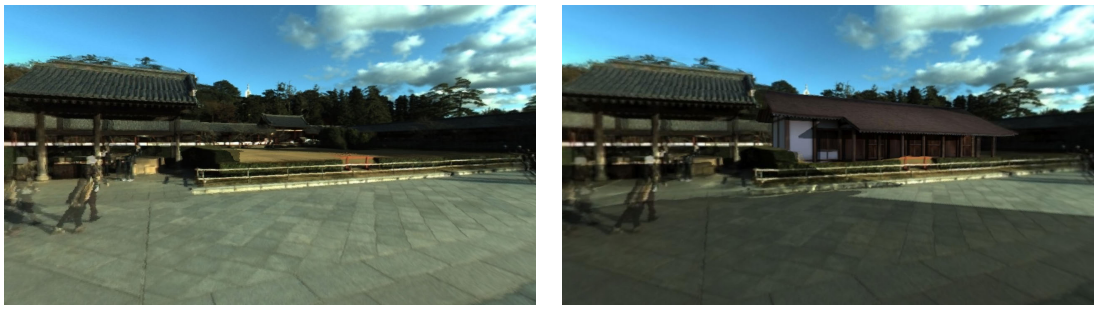
Figure 5.21. Panoramic appearance of MR-world at a structured viewpoint for virtual walk-through. The virtual models and the real environment are not semantically related; they are non-historical contents.



(a)



(b)



(c)

Figure 5.22. Virtual walk-through using 1D structure, whose viewpoints are illustrated in Figure 5.20, with and without superimposition of virtual objects.

ment, it can provide a similar experience to a photorealistic AR. This type of AR system, which displays pre-captured images as the background images for AR as a substitute of the real-world images captured in real-time, is referred to as Indirect AR [WTA11, AOSY13]; however, current Indirect AR systems do not provide changes in user viewpoints. The proposed framework in principle extends Indirect AR to allow for a change in user viewpoint, and offers a novel type of AR application.

## 5.6. Summary

This chapter has described four applications that fully or partially use the proposed imaging and rendering frameworks: 1) an HDR immersive panorama, 2) augmented-immersive panorama, 3) real-time AR, and 4) MR-world exploration. HDR immersive panorama applications can be developed using full spherical HDR images acquired by our imaging framework. We investigated tone-mapping methods of full spherical HDR images using a practical immersive panorama system, and the results show that a tone-mapping method that depends on the user's viewing direction as well as an ordinary tone-mapping yields the best results with respect to critical human factors. A virtual historical tourism application based on an augmented immersive panorama system was demonstrated for over one-thousand people through a public demonstration, and we found that a high-quality user interface, in addition to a photorealistic superimposition of virtual objects, is demanded. We also showed that real-time AR applications can be developed based on the proposed rendering framework. Augmented scenes were successfully generated with high performance using appearance-based free-viewpoint image generation, where the application performance was 10- to 100-times faster than the simple model-based rendering of virtual objects. Based on the implementation described in 4.3, we developed an application providing a change in viewpoint in a 3D space as a straightforward extension of the 2D application, as well as a virtual walk-through application using spherical images captured during walking on the ground.

# Chapter 6

# Conclusions

This thesis has described a novel framework for acquiring real-world scenes, as well as the rendering both of the real and virtual worlds for photorealistic MR-world exploration. To realize a photorealistic and immersive MR scene representation, Chapter 2 illustrated that full spherical HDR images should be acquired through an effective and accurate acquisition of real-world images, and by using rendering methods that can be utilized offline with high-quality illumination through time-consuming methods and a high number of manual operations.

For imaging issues, in particular, it has been difficult to acquire full spherical HDR images from the sky that can be used for virtual globe applications providing a bird's-eye view; therefore, two approaches for capturing full spherical aerial HDR images were described in Chapter 3. The first approach only requires ordinary spherical aerial imaging equipment, which consists of an aerial vehicle and an OMS mounted on the bottom. Spherical images captured by this equipment include missing areas, and therefore must be completed using a statistical model of the sky luminance and radiance. The second approach requires very special hardware configurations, that is, two OMSs mounted on the top and bottom of an aerial vehicle. The two OMSs cover a whole viewing direction, and full spherical images are therefore generated by mutually completing the missing areas appearing in the spherical images captured by each OMS. In addition, the two OMSs capture their own multi-exposure images for generating HDR images. Our experiments demonstrated that the approach using two OMSs generated a notably higher intensity compared with a spherical image completion using a sky

model.

To achieve a photorealistic superimposition of virtual objects onto a real-world virtualization, we propose a novel framework for the photorealistic superimposition of static virtual objects into a virtualized real-world setting, as described in Chapter 4. Offline rendering and a free-viewpoint image generation are combined to preserve the high quality of offline rendering. The computational cost of the online process is highly reduced using pre-generated structured viewpoints and pre-rendering virtual objects onto these viewpoints. We also introduced a practical implementation based on the proposed rendering framework, which inputs spherical images captured using the proposed imaging approach. This framework provides a free configuration of the user's viewpoint on a 2D plane through a VDTM-based free-viewpoint generation technique using pre-generated viewpoints at the grid points. Our implementation demonstrated a high performance during an experiment conducted at a historical site.

Chapter 5 described four categories of applications that fully or partially use the proposed imaging and rendering frameworks: 1) an HDR immersive panorama, 2) augmented immersive panorama, 3) real-time AR, and 4) MR-world exploration. We investigated tone-mapping methods for a practical immersive panorama system using full spherical HDR images acquired using our imaging approach, and the results indicate that a tone-mapping method that depends on the user's viewing direction, as well as an ordinary tone-mapping are promising. A virtual historical tourism application based on an augmented immersive panorama system was demonstrated for over one-thousand people through a public demonstration, and we found that the participants demand a high-quality user interface in addition to a photorealistic superimposition of virtual objects. We also showed that real-time AR applications can be developed based on the proposed rendering framework. In an experiment on a real-time AR application using appearance-based free-viewpoint image generation, augmented scenes were successfully generated at a high performance 10- to 100-times faster than a simple model-based rendering of virtual objects. Two applications for MR-world exploration are also developed: a fly-through application providing a change in viewpoint in a 3D space as a straightforward extension of the implementation described in Section 4.3, and a virtual walk-through application using spherical

images captured during walking on the ground.

Although the proposed framework can be employed in various applications, some issues can be further considered. The proposed imaging approaches do not deal with a control method of the aerial vehicles used for image capturing. To achieve a more effective image acquisition, an automatic vehicle path planning method should be developed. In addition to our imaging framework, such method may be an important issue for 3D reconstruction and free-viewpoint image generation. An important issue of the proposed rendering framework is to use an enormous number of images. Applications using larger environment can be developed by overcoming the problems, such as by using multi-viewpoint image-compression approaches.

# Acknowledgements

This work was completed under the supervision of Professor Naokazu Yokoya and Associate Professor Masayuki Kanbara of the Graduate School of Information Science at Nara Institute of Science and Technology.

The author would like to express his deepest appreciation to Professor Naokazu Yokoya, who offered continuing support and constant encouragement. Without the encouragement, this dissertation would not have been possible. He would particularly like to thank Associate Professor Masayuki Kanbara, who provided him with numerous invaluable comments, suggestions, and ideas. Their extensive and constructive discussions, which would occasionally last for hours, were a driving force for his research.

The author is also deeply grateful to Professor Hirokazu Kato and Associate Professor Tomokazu Sato, who, as members of the thesis committee, offered him insightful comments and suggestions.

The author is also benefited greatly by having the opportunity to meet and work with a number of people during his doctoral work at the Nara Institute of Science and Technology. Secretaries Mio Takahashi, Mina Nakamura, Chika Kijima, and Yumi Ishitani of the Vision and Media Computing Laboratory gave him invaluable support in daily affairs. He would like to express his appreciation to Assistant Professors Norihiko Kawai and Yuta Nakashima of the Vision and Media Computing Laboratory, who gave him constructive comments and suggestions. The author would also like to thank all of the members of the Vision and Media Computing Laboratory of the Graduate School of Information Science at the Nara Institute of Science and Technology for their friendship and support, as well as their cooperation with his experiments.

The author would like to offer his special thanks to Professor Petri Pulli of the

University of Oulu, who provided him with an opportunity to visit the University of Oulu for three months. He appreciates the support of Assistant Professor Goshiro Yamamoto for his research activity and daily life in Oulu. He would also like to express his gratitude to the Japan Student Services Organization, the Japan Society for the Promotion of Science, and the Nara Institute of Science and Technology for their financial support for the visit.

The author would like to express his gratitude to the Japan Society for the Promotion of Science for their financial support through JSPS Research Fellow DC2. He would also like to acknowledge his appreciation for the courtesy extended by Toppan Printing Co., Ltd. for the use of their CG models of Heijo Palace described in Chapter 4.

Finally, the author wants to particularly thank his family for their tireless understanding and support during his Ph.D. studies.

# References


[ABB+01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[ADF+10] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. Google Street View: Capturing the world at street level. *IEEE Computer Magazine*, vol. 43, no. 6, pp. 32–38, 2010.

[AHK+10] Ismail Arai, Maiya Hori, Norihiko Kawai, Yohei Abe, Masahiro Ichikawa, Yusuke Satonaka, Tatsuki Nitta, Tomoyuki Nitta, Harumitsu Fujii, Masaki Mukai, Soichiro Hiromi, Koji Makita, Masayuki Kanbara, Nobuhiko Nishio, and Naokazu Yokoya. Pano UMECHIKA: A crowded underground city panoramic view system. *Proc. Int'l Symp. on Distributed Computing and Artificial Intelligence (DCAI'10)*, pp. 173–180, Valencia, Spain, 2010.

[ALCS03] Kusuma Agusanto, Li Li, Zhu Chuangui, and Ng Wan Sing. Photorealistic rendering for augmented reality using environment illumination. *Proc. Second IEEE/ACM Int'l Symp. on Mixed and Augmented Reality (ISMAR'03)*, pp. 208–216, Tokyo, Japan, 2003.

[AOSY13] Takayuki Akaguma, Fumio Okura, Tomokazu Sato, and Naokazu Yokoya. Mobile AR using pre-captured omnidirectional images. *Proc. ACM SIGGRAPH Asia'13 Symp. on Mobile Graphics and Interactive Applications*, pp. 26:1–26:4, Hong Kong, China, 2013.

[AS02] Shai Avidan and Amnon Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Trans. on Visualization and Computer Graphics*, vol. 4, no. 4, pp. 293–306, 2002.

[ASS+09] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building Rome in a day. *Proc. 12th IEEE Int'l Conf. on Computer Vision (ICCV'09)*, pp. 72–79, Kyoto, Japan, 2009.

[ASSS10] Sameer Agarwal, Noah Snavely, Steven M. Seitz, and Richard Szeliski. Bundle adjustment in the large. *Proc. 11th European Conf. on Computer Vision (ECCV'10)*, pp. 29–42, Crete, Greece, 2010.

[Azu97] Ronald Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[BBM+01] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. Unstructured lumigraph rendering. *Proc. ACM SIGGRAPH'01*, pp. 425–432, Los Angeles, CA, 2001.

[BL06] Andrea Bottino and Aldo Laurentini. What's NEXT? An interactive next best view approach. *Pattern Recognition*, vol. 39, no. 1, pp. 126–132, 2006.

[BMOI08] Atsuhiko Banno, Tomohito Masuda, Takeshi Oishi, and Katsushi Ikeuchi. Flying laser range sensor for large-scale site-modeling and its applications in Bayon digital archival project. *Int'l Journal of Computer Vision*, vol. 78, no. 2, pp. 207–222, 2008.

[Bra97] Richard Bradley. *Rock Art and the Prehistory of Atlantic Europe: Signing the Land*. Psychology Press, New York, NY, 1997.

[But06] Declan Butler. Virtual globes: The web-wide world. *Nature*, vol. 439, no. 7078, pp. 776–778, 2006.

[CCWG88] Michael F. Cohen, Shenchang Eric Chen, John R. Wallace, and Donald P. Greenberg. A progressive refinement approach to fast radiosity

image generation. *Proc. ACM SIGGRAPH'88*, pp. 75–84, Atlanta, GA, 1988.

[CDSHD13] Gaurav Chaurasia, Sylvain Duchene, Olga Sorkine-Hornung, and George Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. on Graphics*, vol. 32, no. 3, pp. 30:1–30:12, 2013.

[Che95] Shenchang Eric Chen. Quicktime VR: An image-based approach to virtual environment navigation. *Proc. ACM SIGGRAPH'95*, pp. 29–38, Los Angeles, CA, 1995.

[Cho09] Aidan Chopra. *Google SketchUp 7 for Dummies*. Wiley, Indianapolis, IN, 2009.

[Cro77] Franklin C. Crow. Shadow algorithms for computer graphics. *Proc. ACM SIGGRAPH'77*, pp. 242–248, San Jose, CA, 1977.

[CSD11] Gaurav Chaurasia, Olga Sorkine, and George Drettakis. Silhouette-aware warping for image-based rendering. *Computer Graphics Forum*, vol. 30, no. 4, pp. 1223–1232, 2011.

[Deb98] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. ACM SIGGRAPH'98*, pp. 189–198, Orlando, FL, 1998.

[DLD12] Abe Davis, Marc Levoy, and Fredo Durand. Unstructured light fields. *Computer Graphics Forum*, vol. 31, no. 2pt1, pp. 305–314, 2012.

[DM97] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *Proc. ACM SIGGRAPH'97*, pp. 369–378, Los Angeles, CA, 1997.

[DT01] Paul Debevec and Chris Tchou. HDR shop. http://www.hdrshop.com/, 2001.

[DTM96] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *Proc. ACM SIGGRAPH'96*, pp. 11–20, New Orleans, LA, 1996.

[DYB98] Paul Debevec, Yizhou Yu, and George Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. *Proc. Ninth Eurographics Workshop on Rendering (EGWR'98)*, pp. 105–116, Vienna, Austria, 1998.

[FB81] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[FB10] Arturo Flores and Serge Belongie. Removing pedestrians from google street view images. *Proc. First IEEE Int'l Workshop on Mobile Vision*, pp. 53–58, San Francisco, CA, 2010.

[FBS06] Jan Fischer, Dirk Bartz, and Wolfgang Straßer. Enhanced visual realism by incorporating camera image effects. *Proc. Fifth IEEE/ACM Int'l Symp. on Mixed and Augmented Reality (ISMAR'06)*, pp. 205–208, Santa Barbara, CA, 2006.

[FGR93] Alain Fournier, Atjeng S. Gunawan, and Chris Romanzin. Common illumination between real and computer generated scenes. *Proc. Graphics Interface'93*, pp. 254–262, Toronto, Canada, 1993.

[FYS+08] Takanori Fukao, Akito Yuzuriha, Takafumi Suzuki, Takeshi Kanzawa, Takashi Oshibuchi, Koichi Osuka, Takashi Kohno, Masahiro Okuyama, Yasuhito Tomoi, and Masaaki Nakadate. Inverse optimal velocity field control of an outdoor blimp robot. *Proc. 17th IFAC World Congress*, pp. 4374–4379, Seoul, Korea, 2008.

[GB08] Payam Ghadirian and Ian D. Bishop. Integration of augmented reality and GIS: A new approach to realistic landscape visualisation. *Landscape and Urban Planning*, vol. 86, pp. 226–232, 2008.

[GCHH03] Simon Gibson, Jon Cook, Toby Howard, and Roger Hubbold. Rapid shadow generation in real-world lighting environments. *Proc. 14th Eurographics Symp. on Rendering (EGSR'03)*, pp. 219–229, Leuven, Belgium, 2003.

[GDS10] Thomas Gierlinger, Daniel Danch, and André Stork. Rendering techniques for mixed reality. *Journal of Real-Time Image Processing*, vol. 5, no. 2, pp. 109–120, 2010.

[GGSC96] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. *Proc. ACM SIGGRAPH'96*, pp. 43–54, New Orleans, LA, 1996.

[GH97] Michael Garland and Paul S. Heckbert. Surface simplification using quadric error metrics. *Proc. ACM SIGGRAPH'97*, pp. 209–216, Los Angeles, CA, 1997.

[Gla89] Andrews Glassner. *An Introduction to Ray Tracing*. Morgan Kaufmann, San Francisco, CA, 1989.

[GM00] Simon Gibson and Alan Murta. Interactive rendering with real-world illumination. *Proc. 11th Eurographics Workshop on Rendering (EGWR'00)*, pp. 365–376, Brno, Czech Republic, 2000.

[GMG+08] Michael A. Goodrich, Bryan S. Morse, Damon Gerhardt, Joseph L. Cooper, Morgan Quigley, Julie A. Adams, and Curtis Humphrey. Supporting wilderness search and rescue using a camera-equipped mini UAV. *Journal of Field Robotics*, vol. 25, no. 1, pp. 89–110, 2008.

[GN03] Michael D. Grossberg and Shree K. Nayar. High dynamic range from multiple images: Which exposures to combine? *Proc. IEEE Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, pp. 1–8, Beijing, China, 2003.

[Gro05] Thorsten Grosch. PanoAR: Interactive augmentation of omni-directional images with consistent lighting. *Proc. Computer Vi-*

*sion/Computer Graphics Collaboration Techniques and Applications (Mirage'05)*, pp. 25–34, Rocquencourt, France, 2005.

[Gro06] Thorsten Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Proc. 2006 Vision Modeling and Visualization (VMV'06)*, pp. 277–284, Aachen, Germany, 2006.

[GRTS12] Lukas Gruber, Thomas Richter-Trummer, and Dieter Schmalstieg. Real-time photometric registration from arbitrary geometry. *Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12)*, pp. 119–128, Atlanta, GA, 2012.

[GWWH03] Nolan Goodnight, Rui Wang, Cliff Woolley, and Greg Humphreys. Interactive time-dependent tone mapping using programmable graphics hardware. *Proc. 14th Eurographics Symp. on Rendering (EGSR'03)*, pp. 26–37, Leuven, Belgium, 2003.

[HDF10] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. *Proc. 2010 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'10)*, pp. 553–560, San Francisco, CA, 2010.

[HDH03] Michael Haller, Stephan Drab, and Werner Hartmann. A real-time shadow approach for an augmented reality application using shadow volumes. *Proc. 10th ACM Symp. on Virtual Reality Software and Technology (VRST'03)*, pp. 56–65, Osaka, Japan, 2003.

[HJD+04] Stanley Herwitz, Lee Johnson, Steve Dunagan, Robert Higgins, Don Sullivan, Jian Zheng, Bradley Lobitz, Joe Leung, Bruce Gallmeyer, Michio Aoyagi, Robert Slye, and James Brass. Imaging from an unmanned aerial vehicle: Agricultural surveillance and decision support. *Computers and Electronics in Agriculture*, vol. 44, pp. 49–61, 2004.

[HJSL04] Emmanuel Hygounenc, Il-Kyun Jung, Philippe Soueres, and Simon Lacroix. The autonomous blimp project of LAAS-CNRS: Achieve-

ments in flight control and terrain mapping. *Int'l Journal of Robotics Research*, vol. 23, no. 4–5, pp. 473–511, 2004.

[HKY09] Maiya Hori, Masayuki Kanbara, and Naokazu Yokoya. MR telepresence system with inertial force sensation using a motion platform and an immersive display. *Proc. IEEE Symp. on 3D User Interfaces (3DUI'09)*, pp. 133–134, Lafayette, LA, 2009.

[HKY10] Maiya Hori, Masayuki Kanbara, and Naokazu Yokoya. Arbitrary stereoscopic view generation using multiple omnidirectional image sequences. *Proc. 20th IAPR Int'l Conf. on Pattern Recognition (ICPR'10)*, pp. 286–289, Istanbul, Turkey, 2010.

[HLHS03] Jean-Marc Hasenfratz, Marc Lapierre, Nicolas Holzschuch, and François X. Sillion. A survey of real-time soft shadows algorithms. *Computer Graphics Forum*, vol. 22, no. 4, pp. 753–774, 2003.

[HS03] Stefan Hrabar and Gaurav S. Sukhatme. Omnidirectional vision for an autonomous helicopter. *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA'03)*, vol. 1, pp. 558–563, Taipei, Taiwan, 2003.

[IHA02] Michal Irani, Tal Hassner, and P. Anandan. What does the scene look like from a scene point? *Proc. Seventh European Conf. on Computer Vision (ECCV'02)*, pp. 883–897, Copenhagen, Denmark, 2002.

[IKH+11] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, and Andrew Davison. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. *Proc. 24th ACM Symp. on User Interface Software and Technology (UIST'11)*, pp. 559–568, Santa Barbara, CA, 2011.

[IKMN04] Norio Igawa, Yasuko Koga, Tomoko Matsuzawa, and Hiroshi Nakamura. Models of sky radiance distribution and sky luminance distribution. *Solar Energy*, vol. 77, pp. 137–157, 2004.

[ISKY04] Sei Ikeda, Tomokazu Sato, Masayuki Kanbara, and Naokazu Yokoya. An immersive telepresence system with a locomotion interface using high-resolution omnidirectional movies. *Proc. 17th Int'l Conf. on Pattern Recognition (ICPR'04)*, vol. 4, pp. 396–399, Cambridge, UK, 2004.

[ISY03] Sei Ikeda, Tomokazu Sato, and Naokazu Yokoya. High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system. *Proc. IEEE Int'l Conf. on Multisensor Fusion and Integration for Intelligent System (MFI'03)*, pp. 155–160, Tokyo, Japan, 2003.

[Jen96] Henrik Wann Jensen. Global illumination using photon maps. *Proc. Seventh Eurographics Workshop on Rendering (EGWR'96)*, pp. 21–30, Porto, Portugal, 1996.

[JLW08] Katrien Jacobs, Celine Loscos, and Greg Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, vol. 28, no. 2, pp. 84–93, 2008.

[JND12] Jan Jachnik, Richard A. Newcombe, and Andrew J. Davison. Real-time surface light-field capture for augmentation of planar specular surfaces. *Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12)*, pp. 91–97, Atlanta, GA, 2012.

[JP11] Michal Jancosek and Tomás Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. *Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, pp. 3121–3128, Colorado Springs, CO, 2011.

[KCSC10] Johannes Kopf, Billy Chen, Richard Szeliski, and Michael Cohen. Street slide: Browsing street level imagery. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'10)*, vol. 29, no. 4, pp. 96:1–96:8, 2010.

[KK12] Peter Kán and Hannes Kaufmann. High-quality reflections, refractions, and caustics in augmented reality and their contribution to

visual coherence. *Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12)*, pp. 99–108, Atlanta, GA, 2012.

[KLS+13] Johannes Kopf, Fabian Langguth, Daniel Scharstein, Richard Szeliski, and Michael Goesele. Image-based rendering in the gradient domain. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH ASIA'13)*, vol. 32, no. 6, pp. 199:1–199:9, 2013.

[KMSY10] Norihiko Kawai, Kotaro Machikita, Tomokazu Sato, and Naokazu Yokoya. Video completion for generating omnidirectional video without invisible areas. *IPSJ Trans. on Computer Vision and Applications*, vol. 2, pp. 200–213, 2010.

[KOKY12] Masaki Kitaura, Fumio Okura, Masayuki Kanbara, and Naokazu Yokoya. Tone mapping for HDR images with dimidiate luminance and spatial distributions of bright and dark regions. *Proc. SPIE Electronic Imaging*, vol. 8292, pp. 829205:1–829205:11, San Francisco, CA, 2012.

[KSA00] Sing Bing Kang, Richard Szeliski, and P. Anandan. The geometry-image representation tradeoff for rendering. *Proc. 2000 IEEE Int'l Conf. on Image Processing (ICIP'00)*, vol. 2, pp. 13–16, Vancouver, BC, 2000.

[KSY09] Norihiko Kawai, Tomokazu Sato, and Naokazu Yokoya. Image inpainting considering brightness change and spatial locality of textures and its evaluation. *Proc. Third Pacific-Rim Symp. on Image and Video Technology (PSIVT'09)*, pp. 271–282, Tokyo, Japan, 2009.

[KTM+10] Martin Knecht, Christoph Traxler, Oliver Mattausch, Werner Purgathofer, and Michael Wimmer. Differential instant radiosity for mixed reality. *Proc. Ninth IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'10)*, pp. 99–107, Seoul, Korea, 2010.

[KTSY10] Hideyuki Kume, Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya. Extrinsic camera parameter estimation using video images

and GPS considering GPS positioning accuracy. *Proc. 20th IAPR Int'l Conf. on Pattern Recognition (ICPR'10)*, pp. 3923–3926, Istanbul, Turkey, 2010.

[KUWS03] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'03)*, vol. 22, no. 3, pp. 319–325, 2003.

[KY02] Masayuki Kanbara and Naokazu Yokoya. Geometric and photometric registration for real-time augmented reality. *Proc. First IEEE/ACM Int'l Symp. on Mixed and Augmented Reality (ISMAR'02)*, pp. 279–280, Darmstadt, Germany, 2002.

[KZP+13] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'13)*, vol. 32, no. 4, pp. 73:1–73:11, 2013.

[LA09] Manolis I. A. Lourakis and Antonis A. Argyros. SBA: A software package for generic sparse bundle adjustment. *ACM Trans. on Mathematical Software*, vol. 36, no. 1, pp. 2:1–2:30, 2009.

[LB12] Philipp Lensing and Wolfgang Broll. Instant indirect illumination for dynamic mixed reality scenes. *Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12)*, pp. 109–118, Atlanta, GA, 2012.

[LH96] Marc Levoy and Pat Hanrahan. Light field rendering. *Proc. ACM SIGGRAPH'96*, pp. 31–42, New Orleans, LA, 1996.

[Lhu11] Maxime Lhuillier. Fusion of GPS and structure-from-motion using constrained bundle adjustments. *Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, pp. 3025–3032, Colorado Springs, CO, 2011.

[LLW04] Ping-Man Lam, Chi-Sing Leung, and Tien-Tsin Wong. A compression method for a massive image data set in image-based rendering.

*Signal Processing: Image Communication*, vol. 19, no. 8, pp. 741–754, 2004.

[MAW+07] Paul Merrell, Amir Akbarzadeh, Liang Wang, Philippos Mordohai, Jan-Michael Frahm, Ruigang Yang, David Nistér, and Marc Pollefeys. Real-time visibility-based fusion of depth maps. *Proc. 11th IEEE Int'l Conf. on Computer Vision (ICCV'07)*, pp. 1–8, Rio de Janeiro, Brazil, 2007.

[MB95] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. *Proc. ACM SIGGRAPH'95*, pp. 39–46, Los Angeles, CA, 1995.

[MCMO10] Iván F Mondragón, Pascual Campoy, Carol Martinez, and Miguel Olivares. Omnidirectional vision applied to unmanned aerial vehicles (UAVs) attitude and heading estimation. *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 809 – 819, 2010.

[MD96] Russell A. Manning and Charles R. Dyer. View morphing. *Proc. ACM SIGGRAPH'96*, pp. 21–30, New Orleans, LA, 1996.

[Min80] Marvin Minsky. Telepresence. *Omni*, vol. 2, no. 9, pp. 45–51, 1980.

[Miy64] Kenro Miyamoto. Fish eye lens. *Journal of the Optical Society of America*, vol. 54, no. 8, pp. 1060–1061, 1964.

[MK94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. on Information Systems*, vol. E77-D, no. 9, pp. 1321–1329, 1994.

[MN99] Tomoo Mitsunaga and Shree K. Nayar. Radiometric self calibration. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'99)*, vol. 1, pp. 374–380, Fort Collins, CO, 1999.

[Moe97] Saied Moezzi. Special issue on immersive telepresence. *IEEE Multimedia*, vol. 4, no. 1, pp. 17–56, 1997.

[MRG03] Marcus Magnor, Prashant Ramanathan, and Bernd Girod. Multi-view coding for image-based rendering using 3-D scene geometry. *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1092–1106, 2003.

[MTUK94] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality virtuality continuum. *Proc. SPIE Telemanipulator and Telepresence Technologies*, vol. 2351, no. 34, pp. 282–292, San Francisco, CA, 1994.

[Nal96] Vic Nalwa. A true omnidirectional viewer. Technical Report BL0115500-960115-01, Bell Laboratories, 1996.

[Nay88] Shree K. Nayar. Sphereo: Determining depth using two specular spheres and a single camera. *Proc. SPIE Optics, Illumination, and Image Sensing for Machine Vision III*, pp. 245–254, San Francisco, CA, 1988.

[Nay97] Shree K. Nayar. Catadioptric omnidirectional camera. *Proc. 1997 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 482–488, San Juan, PR, 1997.

[NCS+09] Masahiko Nagai, Tianen Chen, Ryosuke Shibasaki, Hideo Kumagai, and Afzal Ahmed. UAV-borne 3-D mapping system by multisensor integration. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 701–708, 2009.

[NHIN86] Eihachiro Nakamae, Koichi Harada, Takao Ishizaki, and Tomoyuki Nishita. A montage method: The overlaying of the computer generated images onto a background photograph. *Proc. ACM SIGGRAPH'86*, pp. 207–214, Dallas, TX, 1986.

[NSD94] Jeffry S. Nimeroff, Eero Simoncelli, and Julie Dorsey. Efficient re-rendering of naturally illuminated environments. *Proc. Fifth Eurographics Workshop on Rendering (EGWR'94)*, pp. 359–373, Darmstadt, Germany, 1994.

[NTKH97] Takeshi Naemura, Takahide Takano, Masahide Kaneko, and Hiroshi Harashima. Ray-based creation of photo-realistic virtual world. *Proc. Third Int'l Conf. on Virtual Systems and Multimedia (VSMM'97)*, pp. 59–68, Geneva, Switzerland, 1997.

[OKY10] Fumio Okura, Masayuki Kanbara, and Naokazu Yokoya. Augmented telepresence using autopilot airship and omni-directional camera. *Proc. Ninth IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'10)*, pp. 259–260, Seoul, Korea, 2010.

[OUSY13] Fumio Okura, Yuko Ueda, Tomokazu Sato, and Naokazu Yokoya. Teleoperation of mobile robots by generating augmented free-viewpoint images. *Proc. 2013 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS'13)*, pp. 665–671, Tokyo, Japan, 2013.

[OYTY98] Yoshio Onoe, Kazumasa Yamazawa, Haruo Takemura, and Naokazu Yokoya. Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, vol. 71, no. 2, pp. 154–165, 1998.

[PARJ+06] Ely Carneiro de Paiva, José Raul Azinheira, Josué G. Ramos Jr, Alexandra Moutinho, and Samuel Siqueira Bueno. Project AURORA: Infrastructure and flight control experiments for a robotic airship. *Journal of Field Robotics*, vol. 23, no. 2–3, pp. 201–222, 2006.

[PDG05] Damien Porquet, Jean-Michel Dischler, and Djamchid Ghazanfarpour. Real-time high-quality view-dependent texture mapping using per-pixel visibility. *Proc. Third Int'l Conf. on Computer Graphics and Interactive Techniques in Australasia and South East Asia (GRAPHITE'05)*, pp. 213–220, Dunedin, New Zealand, 2005.

[PH10] Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, Burlington, MA, 2010.

[PHGD+12] Alistair WG Pike, DL Hoffmann, Marcos García-Diez, Paul B Pettitt, Jose Alcolea, Rodrigo De Balbin, César González-Sainz, Carmen de las Heras, Jose A Lasheras, Ramón Montes, and J Zilhäo. U-series dating of paleolithic art in 11 caves in Spain. *Science*, vol. 336, no. 6087, pp. 1409–1413, 2012.

[Pit99] Richard Pito. A solution to the next best view problem for automated surface acquisition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1016–1030, 1999.

[PNF+08] Marc Pollefeys, David Nistér, Jan-Michael Frahm, Amir Akbarzadeh, Philippos Mordohai, Brian Clipp, Chris Engels, David Gallup, S-J Kim, and Paul Merrell. Detailed real-time urban 3D reconstruction from video. *Int'l Journal of Computer Vision*, vol. 78, no. 2, pp. 143–167, 2008.

[Rei02] Erik Reinhard. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools*, vol. 7, no. 1, pp. 45–51, 2002.

[RHD+10] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann, San Francisco, CA, 2010.

[RKB04] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'04)*, vol. 23, no. 3, pp. 309–314, 2004.

[RSSF02] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'02)*, vol. 21, no. 3, pp. 267–276, 2002.

[SCK07] Heung-Yeung Shum, Shing-Chouw Chan, and Sing Bing Kang. *Image-Based Rendering*. Springer, New York, NY, 2007.

[SHK+05] Imari Sato, Morihiro Hayashida, Fumiyo Kai, Yoichi Sato, and Katsushi Ikeuchi. Fast image synthesis of virtual objects in a real scene with natural shadings. *Systems and Computers in Japan*, vol. 36, no. 14, pp. 102–111, 2005.

[SIY04] Tomokazu Sato, Sei Ikeda, and Naokazu Yokoya. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. *Proc. European Conf. on Computer Vision (ECCV'04)*, vol. 2, pp. 326–340, Prague, Czech Republic, 2004.

[SKC03] Heung-Yeung Shum, Sing Bing Kang, and Shing-Chow Chan. Survey of image-based representations and compression techniques. *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1020–1037, 2003.

[SKG+12] Sudipta N Sinha, Johannes Kopf, Michael Goesele, Daniel Scharstein, and Richard Szeliski. Image-based rendering for scenes with reflections. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'12)*, vol. 31, no. 4, pp. 100:1–100:10, 2012.

[SSI02] Imari Sato, Yoichi Sato, and Katsushi Ikeuchi. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Trans. on Visualization and Computer Graphics*, vol. 5, no. 1, pp. 1077–2626, 2002.

[SSI03] Imari Sato, Yoichi Sato, and Katsushi Ikeuchi. Illumination from shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 290–300, 2003.

[SSS06] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'06)*, vol. 25, no. 3, pp. 835–846, 2006.

[ST94] Jianbo Shi and Carlo Tomasi. Good features to track. *Proc. 1994 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593–600, Seattle, WA, 1994.

[Ste92] Jonathan Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, vol. 42, no. 4, pp. 73–93, 1992.

[STJ+04] Jessi Stumpfel, Chris Tchou, Andrew Jones, Tim Hawkins, Andreas Wenger, and Paul Debevec. Direct HDR capture of the sun and sky. *Proc. Third Int'l Conf. on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa (AFRIGRAPH'04)*, pp. 145–149, Stellenbosch, South Africa, 2004.

[STYY00] Jun Shimamura, Haruo Takemura, Naokazu Yokoya, and Kazumasa Yamazawa. Construction and presentation of a virtual environment using panoramic stereo images of a real scene and computer graphics models. *Proc. 15th Int'l Conf. on Pattern Recognition (ICPR'00)*, vol. 4, pp. 463–467, Barcelona, Spain, 2000.

[Sut65] Ivan E Sutherland. The ultimate display. *Proc. IFIP Congress*, vol. 65, no. 2, pp. 506–508, 1965.

[Sut68] Ivan E Sutherland. A head-mounted three dimensional display. *Proc. Fall Joint Computer Conf.*, vol. 33, pp. 757–764, 1968.

[SY10] Tomokazu Sato and Naokazu Yokoya. Efficient hundreds-baseline stereo by counting interest points for moving omni-directional multi-camera system. *Journal of Visual Communication and Image Representation*, vol. 21, no. 5–6, pp. 416–426, 2010.

[TKIS00] Takuji Takahashi, Hiroshi Kawasaki, Katsushi Ikeuchi, and Masao Sakauchi. Arbitrary view position and direction rendering for large-scale scenes. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, vol. 2, pp. 2296–2303, Hilton Head, SC, 2000.

[TMHF99] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment—a modern synthesis. *Proc. Int'l Workshop on Vision Algorithms*, pp. 298–372, Corfu, Greece, 1999.

[UCK+04] Matthew Uyttendaele, Antonio Criminisi, Sing Bing Kang, Simon Winder, Richard Szeliski, and Richard Hartley. Image-based interactive exploration of real-world environments. *IEEE Computer Graphics and Applications*, vol. 24, no. 3, pp. 52–63, 2004.

[VCL+11] Peter Vangorp, Gaurav Chaurasia, Pierre-Yves Laffont, Roland W Fleming, and George Drettakis. Perception of visual artifacts in image-based rendering of façades. *Computer Graphics Forum*, vol. 30, no. 4, pp. 1241–1250, 2011.

[Vin07] Luc Vincent. Taking online maps down to street level. *IEEE Computer Magazine*, vol. 40, no. 12, pp. 118–120, 2007.

[VRC+13] Peter Vangorp, Christian Richardt, Emily A. Cooper, Gaurav Chaurasia, Martin S. Banks, and George Drettakis. Perception of perspective distortions in image-based rendering. *ACM Trans. on Graphics (Proc. ACM SIGGRAPH'13)*, vol. 32, no. 4, pp. 58:1–58:11, 2013.

[WACS11] Changchang Wu, Sameer Agarwal, Brian Curless, and Steven M. Seitz. Multicore bundle adjustment. *Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, pp. 3057–3064, Colorado Springs, CO, 2011.

[War94] Greg J. Ward. The RADIANCE lighting simulation and rendering system. *Proc. ACM SIGGRAPH'94*, pp. 459–472, Orlando, FL, 1994.

[War03] Greg J. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17–30, 2003.

[Wol08] Robert M. Wolk. Utilizing Google Earth and Google Sketchup to visualize wind farms. *Proc. 2008 IEEE Int'l Symp. on Technology and Society (ISTAS'08)*, pp. 1–8, Fredericton, NB, 2008.

[WPF90] Andrew Woo, Pierre Poulin, and Alain Fournier. A survey of shadow algorithms. *IEEE Computer Graphics and Applications*, vol. 10, no. 6, pp. 13–32, 1990.

[WTA11] Jason Wither, Yun-Ta Tsai, and Ronald Azuma. Indirect augmented reality. *Computers and Graphics*, vol. 35, no. 4, pp. 810–822, 2011.

[Wu13] Changchang Wu. Towards linear-time incremental structure from motion. *Proc. 2013 Int'l Conf. on 3D Vision (3DV'13)*, pp. 127–134, Seattle, WA, 2013.

[XBF+09] Hanwei Xu, Rami Badawi, Xiaohu Fan, Jiayong Ren, and Zhiqiang Zhang. Research for 3D visualization of digital city based on SketchUp and ArcGIS. *Proc. SPIE Int'l Symp. on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining*, vol. 7492, pp. 74920Z:1–74920Z:12, San Francisco, CA, 2009.

[XT97] Yalin Xiong and Kenneth Turkowski. Creating image-based VR using a self-calibrating fisheye lens. *Proc. 1997 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 237–243, San Juan, PR, 1997.

[Yag99] Yasushi Yagi. Omnidirectional sensing and its applications. *IEICE Trans. on Information and Systems*, vol. 82, no. 3, pp. 568–579, 1999.

[YISY06] Yuji Yokochi, Sei Ikeda, Tomokazu Sato, and Naokazu Yokoya. Extrinsic camera parameter estimation based-on feature tracking and GPS data. *Proc. Seventh Asian Conf. on Computer Vision (ACCV'06)*, vol. 1, pp. 369–378, Hyderabad, India, 2006.

[YK90] Yasushi Yagi and Shinjiro Kawato. Panorama scene analysis with conic projection. *Proc. 1990 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS'90)*, pp. 181–187, Ibaraki, Japan, 1990.

[YYY93] Kazumasa Yamazawa, Yasushi Yagi, and Masahiko Yachida. Omnidirectional imaging with hyperboloidal projection. *Proc. 1993*

*IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS'93)*, vol. 2, pp. 1029–1034, Tokyo, Japan, 1993.

[ZC04] Cha Zhang and Tsuhan Chen. A survey on image-based rendering–representation, sampling and compression. *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[ZFPW03] Assaf Zomet, Doron Feldman, Shmuel Peleg, and Daphna Weinshall. Mosaicing new views: The crossed-slits projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 741–754, 2003.

# List of Publications

## Journal Papers

1. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Augmented telepresence using recorded aerial omnidirectional videos captured from unmanned airship. *Trans. of Virtual Reality Society of Japan*, Vol. 16, No. 2, pp. 127–138, Jun 2011 [VRSJ Outstanding Paper Award] (in Japanese). (Chapter 3, Chapter 4, Chapter 5)

2. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Aerial HDR video generation of full spherical views using two omnidirectional camera units. *Trans. of Virtual Reality Society of Japan*, Vol. 17, No. 3, pp. 139–149, Sep 2012 (in Japanese). (Chapter 3).

3. Fumio Okura, Yuko Ueda, Tomokazu Sato, Naokazu Yokoya. Free-viewpoint mobile robot teleoperation interface using view-dependent geometry and texture. *ITE Trans. on Media Technology and Applications*, Vol. 2, No. 1, pp. 82–93, Jan 2014.

## International Conferences

1. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Augmented telepresence from the sky: AR using autopilot airship and omni-directional camera. *Proc. 3rd Korea-Japan Workshop on Mixed Reality (KJMR'10)*, pp. 190–200, Apr 2010. (Chapter 3).

2. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Augmented telepres-

ence using autopilot airship and omni-directional camera. *Proc. 9th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'10)*, pp. 259–260, Oct 2010. (Chapter 3).

3. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Fly-through Heijo Palace Site: Augmented telepresence using aerial omnidirectional videos. *Proc. ACM SIGGRAPH'11 Posters*, Article No. 78, Aug 2011. (Chapter 3, Chapter 4, Chapter 5).

4. Masaki Kitaura, Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Tone mapping for HDR images with dimidiate luminance and spatial distributions of bright and dark regions. *Proc. SPIE Electronic Imaging*, Vol. 8292, pp. 829205-01–829205-11, Jan 2012. (Chapter 5).

5. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Full spherical high dynamic range imaging from the sky. *Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME'12)*, pp. 325–332, Jul 2012. (Chapter 3).

6. Fumio Okura. Spacetime freeview generation using image-based rendering, relighting, and augmented telepresence. *Proc. ACM Multimedia (MM'12) Doctoral Symposium*, pp. 1437–1440, Oct 2012. (Chapter 3).

7. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Fly-through Heijo Palace Site: Historical tourism system using augmented telepresence. *Proc. ACM Multimedia (MM'12) Technical Demo*, pp. 1283–1284, Oct 2012. (Chapter 3, Chapter 4, Chapter 5).

8. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Interactive exploration of augmented aerial scenes with free-viewpoint image generation from pre-rendered images. *Proc. 12th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'13)*, pp. 279–280, Oct 2013. (Chapter 4).

9. Fumio Okura, Yuko Ueda, Tomokazu Sato, Naokazu Yokoya. Teleoperation of mobile robots by generating augmented free-viewpoint images. *Proc. 2013 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS'13)*, pp. 665–671, Nov 2013.

10. Takayuki Akaguma, Fumio Okura, Tomokazu Sato, Naokazu Yokoya. Mobile AR using pre-captured omnidirectional images. *Proc. ACM SIGGRAPH Asia 2013 Symp. on Mobile Graphics and Interactive Applications*, Article No. 26, Nov 2013.

# Domestic Conferences

1. Naoko Nitta, Yukifumi Okada, Nozomu Kasuya, Yusuke Utsuno, Makoto Fujigaki, Shinnosuke Tokumoto, Kenichiro Fuji, Fumio Ogawa, Toru Kawasaki, Takuma Maruyama, Fumio Okura. PRMU Algorithm Contest 2009 : "Find Clones!" summary report and prize winning algorithms. *Technical Report of IEICE*, PRMU2009-155, Dec 2009 (in Japanese).

2. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Aerial imaging system by automatic control of unmanned airship. *Proc. IEICE General Conference 2010*, D-12-84, Mar 2010 (in Japanese).

3. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Augmented telepresence from the sky: Augmented reality using autopilot airship and omnidirectional camera. *Proc. Meeting on Image Recognition and Understanding (MIRU'10)*, pp. 1183–1189, Jul 2010 (in Japanese). (Chapter 3).

4. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Fly-through augmented telepresence using recorded aerial omnidirectional videos. *Proc. 15th Virtual Reality Society of Japan Annual Conference*, pp. 394–397, Sep 2010 (in Japanese). (Chapter 3, Chapter 4, Chapter 5).

5. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Fly-through MR Heijokyo: Augmented telepresence using recorded aerial omnidirectional videos captured from unmanned airship. *Technical Report of IEICE*, MVE2010-58, Oct 2010 (in Japanese). (Chapter 3, Chapter 4, Chapter 5).

6. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Geometric and photometric registration for augmented telepresence using recorded aerial omnidirectional videos. *Proc. Meeting on Image Recognition and Understanding*

*(MIRU'11)*, pp. 1177–1184, Jul 2011 (in Japanese). (Chapter 3, Chapter 4, Chapter 5).

7. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Aerial HDR imaging of complete spherical images using two omnidirectional cameras. *Proc. Forum on Information Technology (FIT'11)*, Vol. 3, pp. 187–190, Sep 2011 (in Japanese). (Chapter 3).

8. Yoshiki Tanaka, Fumio Okura, Maiya Hori, Masayuki Kanbara, Naokazu Yokoya. Omnidirectional image acquisition support system using recommendation degree map for telepresence with image-based rendering. *Proc. 16th Virtual Reality Society of Japan Annual Conference*, pp. 640–641, Sep 2011 (in Japanese).

9. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Aerial HDR imaging of full spherical views without missing areas using two omnidirectional cameras. *Technical Report of IEICE*, MVE2011-80, Jan 2012 [MVE Award] (in Japanese). (Chapter 3).

10. Masaki Kitaura, Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Tone-mapping for HDR images divided with dimidiate luminance and spatial distributions of bright and dark regions. *Technical Report of IEICE*, MVE2011-74, Jan 2012 (in Japanese). (Chapter 5).

11. Yoshiki Tanaka, Fumio Okura, Maiya Hori, Masayuki Kanbara, Naokazu Yokoya. Image acquisition support system using recommendation degree map for telepresence with image-based rendering. *Technical Report of IEICE*, MVE2011-128, Mar 2012 (in Japanese).

12. Masaki Kitaura, Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Tone-mapping using region segmentation for HDR images with dimidiate luminance and spatial distributions. *Proc. Meeting on Image Recognition and Understanding (MIRU'12)*, Aug 2012 (in Japanese). (Chapter 5).

13. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Displaying spherical aerial HDR images for telepresence system. *Technical Report of IEICE*, MVE2012-42, Sep 2012 (in Japanese). (Chapter 5).

14. Yuko Ueda, Fumio Okura, Tomokazu Sato, Naokazu Yokoya. Teleoperation interface of mobile robots providing freely configurable viewpoint. *Proc. 2012 Kansai Joint Convention of Institutes of Electrical Engineering*, pp. 463–464, Dec 2012 (in Japanese).

15. Yuko Ueda, Fumio Okura, Tomokazu Sato, Naokazu Yokoya. Mobile robot control interface using augmented free-viewpoint image generation. *Technical Report of IEICE*, MVE2012-73, Jan 2013 (in Japanese).

16. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya. Photorealistic superimposition of virtual objects onto virtualized real-world using pre-rendering and free-viewpoint image generation. *Proc. Meeting on Image Recognition and Understanding (MIRU'13)*, SS4-29, Jul 2013 (in Japanese). (Chapter 4).

17. Takayuki Akaguma, Fumio Okura, Tomokazu Sato, Naokazu Yokoya. Mobile AR using pre-rendered images. *Proc. 18th Virtual Reality Society of Japan Annual Conference*, pp. 513–516, Sep 2013 (in Japanese).

18. Fumio Okura, Masayuki Kanbara, Naokazu Yokoya.Photorealistic augmented reality based on free-viewpoint image generation using pre-rendered images. *Technical Report of VRSJ*, Vol. 18, No. CS-3, pp. 11–16, Sep 2013 (in Japanese). (Chapter 4).

19. Fumio Okura, Yuko Ueda, Tomokazu Sato, Naokazu Yokoya. Teleoperation interface of mobile robots using free-viewpoint image generation. *Proc. 2013 Kansai Joint Convention of Institutes of Electrical Engineering*, pp. 456–457, Nov 2013 (in Japanese).

20. Takayuki Akaguma, Fumio Okura, Tomokazu Sato, Naokazu Yokoya. AR using pre-captured images considering illumination of real scene. *Technical Report of IEICE*, PRMU2013-161, Feb 2014 (in Japanese).

21. Naoya Inoue, Norihiko Kawai, Tomokazu Sato, Fumio Okura, Yuta Nakashima, Naokazu Yokoya. Removal of moving objects from omnidirectional video taken by a moving camera. *Proc. IEICE General Conference 2014*, D-11-43, Mar 2014 (in Japanese).

# Awards

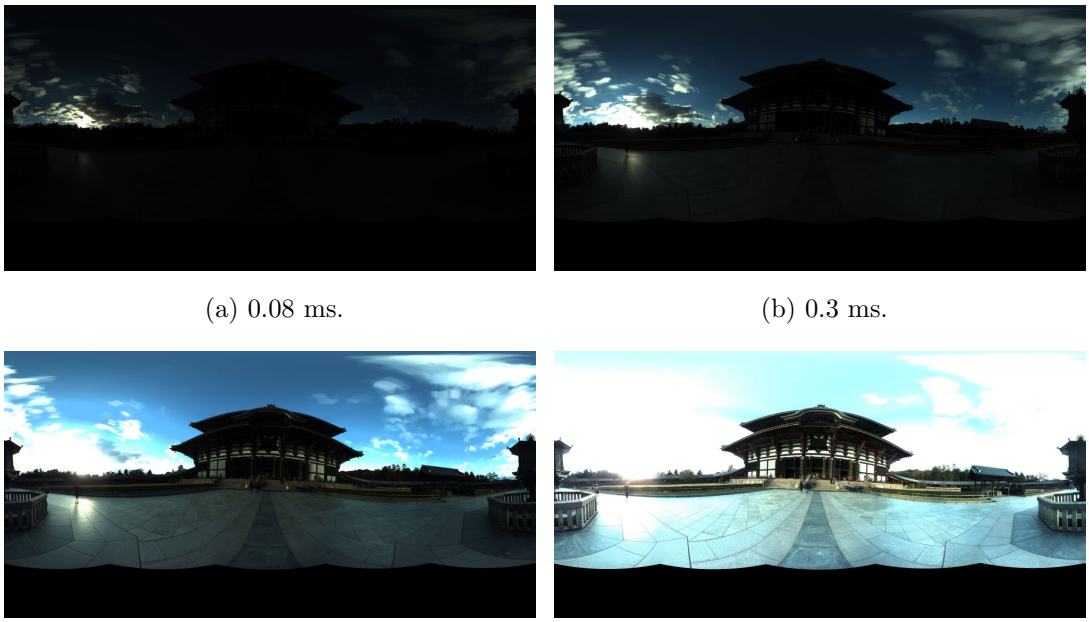
1. Jury's Special Award, IEICE PRMU Algorithm Contest 2009, Sep 2009.
2. 2011 NAIST Top Scholarship Program, Jun 2011.
3. MVE Award, IEICE Technical Group on Multimedia and Virtual Environment (MVE), Sep 2012.
4. VRSJ Outstanding Paper Award, The Virtual Reality Society of Japan, Sep 2012.

# Appendix A

# Spherical HDR Imaging on the Ground Using OMS

Spherical HDR images can be acquired simply from the ground using an OMS, although there seems to be a lack of knowledge accumulated for such imaging methods. In the same fashion as HDR imaging using ordinary monocular cameras, multi-exposure images are first captured using an OMS. Unless capturing in motion is particularly required, the OMS should be mounted on a tripod to avoid a misalignment of the multi-exposure images. The auto exposure determination used for our aerial HDR imaging described in Section 3.3.2 can also be employed for multi-exposure imaging on the ground. The pedestrian subtraction method [AHK+10] can be applied for each multi-exposure image sequence captured at a fixed position. HDR images are generated from a set of multi-exposure images using the method described by Devebec and Malik [DM97].

Figure 2.5 in Section 2.1.3 shows an example of an HDR spherical image captured in an underground city using the method mentioned above after pedestrian removal [AHK+10]. Through this same method, outdoor scenes can also be captured as spherical HDR images. Figure A.1 shows multi-exposure images taken using an auto exposure determination and pedestrian removal that were acquired in an outdoor environment. A spherical HDR image was generated from the multi-exposure images, as shown in Figure A.2.



(a) 0.08 ms. (b) 0.3 ms.

(c) 1.2 ms. (d) 4.8 ms.

Figure A.1. Multi-exposure images whose exposure values are determined using the method described in Section 3.3.2. Pedestrians in the images were removed using the method described in [AHK+10].

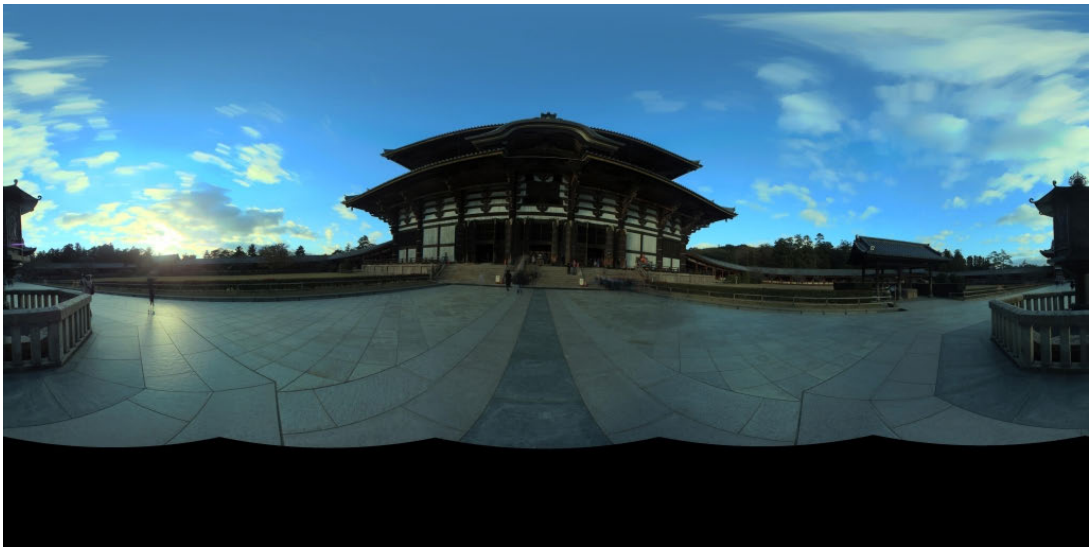


Figure A.2. A spherical HDR image captured in outdoor environment.