

Aerial Full Spherical HDR Imaging and Display

Fumio Okura · Masayuki Kanbara · Naokazu Yokoya

Received: 17 June 2013 / Accepted: 7 August 2014

Abstract This paper describes a framework for aerial imaging of high dynamic range (HDR) scenes for use in virtual reality applications, such as immersive panorama applications and photorealistic superimposition of virtual objects using image-based lighting. We propose a complete and practical system to acquire full spherical HDR images from the sky, using two omnidirectional cameras mounted above and below an unmanned aircraft. The HDR images are generated by combining multiple omnidirectional images captured with different exposures controlled automatically. Our system consists of methods for image completion, alignment, and color correction, as well as a novel approach for automatic exposure control, which selects optimal exposure so as to avoid banding artifacts. Experimental results indicated that our system generated better spherical images compared to an ordinary spherical image completion system in terms of naturalness and accuracy. In addition to proposing an imaging method, we have carried out an experiment about display methods for aerial HDR immersive panoramas utilizing spherical images acquired by the proposed system. The experiment demonstrated HDR imaging is beneficial to immersive panorama using an HMD, in addition to ordinary uses of HDR images.

Keywords Omnidirectional camera · High dynamic range image · Immersive panorama · Image-based lighting · Tone-mapping

F. Okura · M. Kanbara · N. Yokoya
Graduate School of Information Science, Nara Institute of
Science and Technology (NAIST)
Takayama-cho, 8916-5, Ikoma, Nara, Japan
Tel.: +81-743-72-5296
Fax: +81-743-72-5299
E-mail: {fumio-o,kanbara,yokoya}@is.naist.jp

1 Introduction

This paper describes a framework for acquiring full spherical high dynamic range (HDR) aerial images. Whereas traditional studies of image-based lighting (IBL) [Debevec(1998)] have focused on methods for acquiring spherical HDR images as light probes [Reinhard et al(2010)], in recent years, improvements in imaging technology have made acquisition much easier, allowing capture of fully immersive 360° panoramas. The first popular application of this type of immersive imaging was Apple's QuickTime VR [Chen(1995)], which synthesized spherical images from images taken at different rotations around an approximate centerpoint. Today, immersive panorama applications can capture spherical images/videos using omnidirectional cameras, and convert them to view direction-dependent perspective images in real-time. This technique has become crucial to popular, web-based walk-through applications such as Google Street View [Anguelov et al(2010)], as well as emerging immersive systems that use head-mounted displays (HMDs) and head trackers [Onoe et al(1998)]. While most of these systems use eight-bit low dynamic range (LDR) images, it is clear that HDR imaging can significantly improve the fidelity of scenes with high dynamic range, including most outdoor scenes. Recently, work on augmented reality (AR) visualization of cultural heritage [Okura et al(2011)] has provided users with panoramic aerial views overlaid with historical computer graphics (CG) models (of lost architectural structures, e.g.) rendered using image based lighting (IBL) techniques. The success of these systems in conveying realistic historical context offers a glimpse of the larger potential for HDR panoramic imaging.

This paper proposes a novel method for acquiring spherical HDR images from the sky. These images are intended for use in immersive and/or augmented panoramas supporting user fly-through. Our imaging framework also can be used to generate input images for more advancing display methods, such as multi-perspective panoramas [Zomet et al(2003), Kopf et al(2010)] and free-viewpoint image generation using image-based rendering [Debevec et al(1996), Shum et al(2003)]. In addition to proposing an imaging method, we have carried out an experiment about display methods for aerial HDR immersive panoramas utilizing spherical images acquired by the proposed framework. Through the experiment, we have investigated the advantages and features of HDR imaging for immersive panoramas. The contribution of our study is summarized as follows:

- We propose a complete and practical system to capture full spherical aerial HDR images, which includes a novel approach for automatic exposure control.
- An investigation of display methods demonstrates HDR imaging is beneficial to immersive panorama using an HMD, in addition to ordinary uses (i.e., photography and IBL).

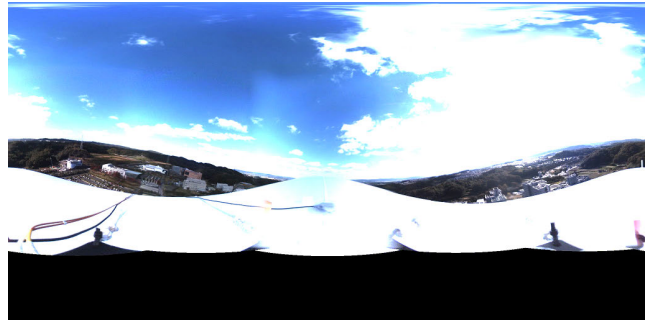
The remainder of this paper is organized as follows. Section 2 describes the difficulties inherent in capturing spherical HDR images from aerial platforms. Section 3 and 4 present our system for acquiring full spherical HDR images using an aerial vehicle and omnidirectional cameras, as well as some experimental results for practical imaging contexts. Section 5 presents and evaluates methods for displaying aerial spherical HDR images in immersive panorama applications. Section 6 provides concluding remarks.

2 Problems in acquiring aerial spherical HDR images

The proposed aerial imaging system resolves two problems that commonly occur when spherical images are acquired using omnidirectional cameras: 1) **the occurrence of missing areas**, and 2) **deficiency in dynamic range**. In this section, we briefly look at each of these problems, along with related work by other researchers, and present our basic approach to solution.

2.1 Occurrence of missing areas

Missing areas, as seen in Figs. 1(a) and 1(b), appear in all single-capture spherical images. They are caused



(a) Image captured above aircraft.



(b) Image captured beneath aircraft.



(c) Full spherical image.

Fig. 1 Full spherical HDR image generation using two omnidirectional cameras.

by limitations in the field of view of the omnidirectional camera and/or occlusion of the field of view by hardware. In the case of Fig. 1, the primary occlusion is the aircraft on which the camera is mounted. Such areas subtract from the immersive value of the image and disqualify it from use in IBL rendering systems.

Existing approaches to filling in such missing areas can be classified into two categories—namely, filling without observed intensities and filling with observed intensities:

1. **Filling without observed intensities.** [Okura et al(2011)] filled missing areas in the upper hemisphere of spherical videos with the aid of a statistical “sky model” of various weather conditions, whereas [Kawai et al(2009)] filled them from the image itself, using inpainting techniques.

2. **Filling with observed intensities.** In a study by [Kawai et al(2010)], missing areas in the ground portion of spherical video sequences were filled by estimating the intensities of the missing area based on the intensities of images of the same location taken from other viewpoints.

The objective of the former of these categories is merely to patch over missing areas in an acceptable manner. The results may differ greatly in intensity from the real objects at the time of capture, or virtual objects rendered using IBL. Our proposed approach falls in the latter category, in which the real intensities are observed. Specifically, we use two omnidirectional cameras, mounted above and below an unmanned aircraft, to capture the entire scene, as shown in Fig. 1(c).

2.2 Deficiency in dynamic range

Proper IBL rendering requires unsaturated spherical images [Debevec and Malik(1997)]. Unfortunately, the dynamic range of many daytime, outdoor scenes is too high to capture using standard eight-bit cameras, as the sun is approximately 2^{17} times brighter than darker areas of sky or cloud cover [Stumpfel et al(2004)]. An HDR imaging method using LDR images taken at multiple exposure values (multi-exposure images), originally proposed by [Debevec and Malik(1997)], is now widely used to capture scenes below saturation using commodity camera hardware. We employ this technique to generate our spherical HDR images.

When multi-exposure images are captured in motion, misalignment will naturally occur. Although solutions to this problem have been proposed (e.g., [Ward(2003)]), large misalignments remain difficult to fix. For this reason, it is best to use fewer multi-exposure images to generate HDR images, and to ensure that camera velocity and frame rate remain constant.

In generating HDR images from a small number of multi-exposure images, particular care must be taken to quantize the luminance. When there are high luminance objects like the sun within the field of view, exposure times must be short. This, in turn, leaves fewer gradations for low luminance objects, which tend to cover most the view area. To adjust for this discrepancy, we employ an automatic exposure control method to accommodate illumination changes while capturing an image sequence. Specifically, exposure times are adapted to the luminance distribution of the scene, based on an HDR histogram generated from multi-exposure images.

An omnidirectional camera mounted atop the aircraft mainly captures sun and sky, which are too bright

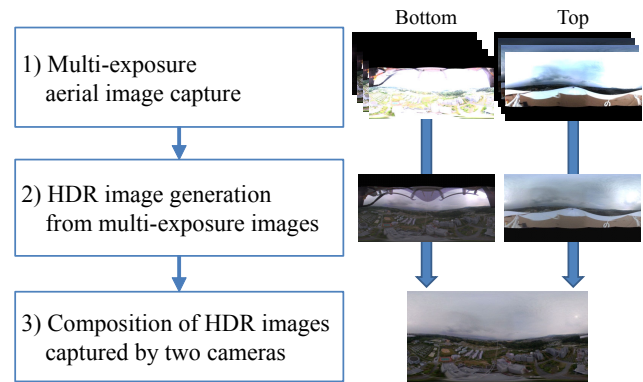


Fig. 2 Overview of aerial full spherical HDR imaging.

to capture without saturation even at the shortest exposure time. To address this problem, neutral density (ND) filters are attached to the camera to overcome, as was done in related work by [Stumpfel et al(2004)].

3 Acquisition of full spherical HDR images

3.1 Overview

Our proposed system generates full spherical HDR images from multi-exposure images captured with two omnidirectional cameras mounted above and below an unmanned aircraft. As depicted in Fig. 2, the procedure is divided into three stages:

1. **Multi-exposure aerial image capture.** Multi-exposure images are captured using two omnidirectional cameras mounted above and below an unmanned aircraft. Since the camera above the vehicle mainly captures the sun and sky, it has ND filters attached. Exposure times are controlled automatically so as to capture high and low luminance images effectively.
2. **HDR image generation from multi-exposure images.** HDR spherical images are generated from the multi-exposure images captured above and below the aircraft. Large misalignments are corrected by estimating changes in 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 in stage two. Because the precise geometric relationship between images from above and below the aircraft is difficult to fix (due to the slight deformability of the unmanned aircraft), we align the images by estimating the relative rotation from one camera to the other for every frame. We also correct chromatic

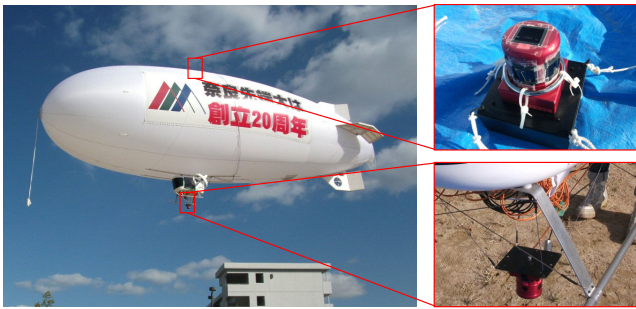


Fig. 3 Unmanned airship and omnidirectional cameras. Two cameras are mounted above and below the aircraft. ND filters are attached to the camera above. A GPS antenna is also mounted atop the aircraft.

changes that may occur in images captured under ND filtering.

3.2 Multi-exposure aerial image capture

3.2.1 Configuration of the aerial imaging system

We used a 12-meter remote-controlled unmanned airship (Fig. 3) as our aerial imaging platform. The aircraft was equipped with a differential GPS (Hitachi Zosen Corp., P4-GPS) for acquiring the locations at which images were captured. Two omnidirectional multi-camera systems (Point Grey Research Ladybug2, see Table 1) were mounted above and below the aircraft. The cameras were connected to a laptop PC for time-stamped storage. ND filters (Fujifilm Corp., ND 2.0) that block all but 1% of visible light were attached to the camera above the vehicle. Figs. 1(a) and 1(b) show panoramic images after removal of limb darkening and geometric transformation [Ikeda et al(2003)]. Note that the amount of limb darkening differs between images captured with and without ND filters.

Although our imaging framework does not require a specific kind of aircraft for mounting the camera, unmanned airships, which were used in our imaging equipment, are beneficial to the imaging for aerial immersive panoramas. They have recently attracted attention for low-height aerial imaging [Herwitz et al(2004), Goodrich et al(2008)], which can acquire higher resolution images. In addition, low-speed vehicles, including airships, are beneficial to multi-exposure imaging because small misalignments occur on the images. However, owing to the deformability of the body of the aircraft, it is difficult to fix the cameras on it; thus, we overcome a problem due to the deformability during the composition of two HDR images.

3.2.2 Automatic exposure control

The exposure times used to capture multi-exposure images are automatically determined from the intensity of previously captured images. Many still cameras provide an auto-bracketing function for capturing multi-exposure images. They determine a proper exposure value with standard auto-exposure controls, and then use fixed multiples of this as the remaining exposure values [Kang et al(2003)]. [Grossberg and Nayar(2003)] proposed a method of determining unfixed exposure sets from the dynamic range of the scene. Our approach, unlike existing approaches, utilizes rich information of the scene to determine a more appropriate exposure set; i.e., apart from the dynamic range, our approach also uses an HDR histogram of the scene, which is acquired by the last multi-exposure image set to reduce the side-effects of large quantization steps on human vision, such as banding (i.e., posterization) resulting from too few multi-exposure images. Note that our approach optimizes the set of exposure times, whereas other modern studies have sought to minimize signal-to-noise ratio (SNR) [Hasinoff et al(2010)]. Such minimization could also be integrated into our approach, as a means of improving the appearance of HDR images, particularly those captured in dark environments.

New exposure times $s_{new_1}, \dots, s_{new_n}$ are determined by multi-exposure images captured using old exposures $s_{old_1}, \dots, s_{old_n}$. In practice, we set $n = 4$ as the number of exposures in our experiment to match the number of registers available in the Ladybug2 cameras for storing exposure times.

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

$$s_{new_1} = \begin{cases} 0.5s_{old_1} & (L_1 = MAX_IN) \\ \frac{MAX_IN + \theta_{sh}}{2L_1} s_{old_1} & (L_1 \leq \theta_{sh}) \\ s_{old_1} & (\text{otherwise}), \end{cases} \quad (1)$$

Table 1 Ladybug2 specifications.

Item	Specification
Number of cameras	Six (five in horizontal ring, and one on top)
Resolution	1024×768 (for each CCD)
FOV	More than 75% of the full sphere
CCD	Sony ICX204AK
Frame rate	Maximum 30fps
Number of shutter registers	Four (to capture multi-exposure images)

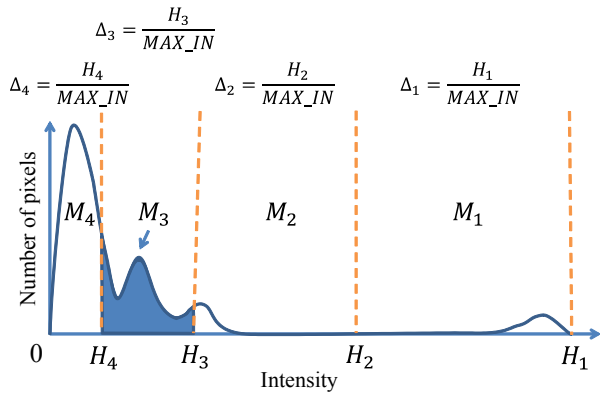


Fig. 4 Illustrative example of HDR histogram for $n = 4$, where Δ_i denotes quantization steps and H_i is the maximum intensity that can be captured using exposure time s_{new_i} . Further, M_3 is the number of pixels in the filled area of the histogram, with the other M_i determined in a similar fashion.

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 eight-bit images). Further, θ_{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 large quantization steps, such as banding, can appear in the generated HDR image. Other exposure times $s_{new_2}, \dots, s_{new_n}$ are therefore determined to reduce such effects. To estimate exposure times that have minimum quantization steps, an HDR histogram is calculated from the captured images, as illustrated in Fig. 4. H_i is the maximum intensity captured using the i -th shortest ($i \geq 2$) exposure time s_{new_i} in the HDR histogram. Number of pixels M_i that belongs to histogram bins between H_{i+1} and H_i (see Fig. 4), is computed using H_i as

$$M_i = \sum_{k=\mathcal{H}}^{H_i} j_k, \quad (2)$$

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

$$\mathcal{H} = \begin{cases} 0 & (i = n) \\ H_{i+1} & (\text{otherwise}). \end{cases} \quad (3)$$

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

$$\Delta_i = \frac{H_i}{MAX_IN}. \quad (4)$$

Finally, s_{new_i} ($i \geq 2$) is determined by varying s_{new_1} to minimize the energy function E_s , where E_s is the sum of the products of the quantization step Δ_i and

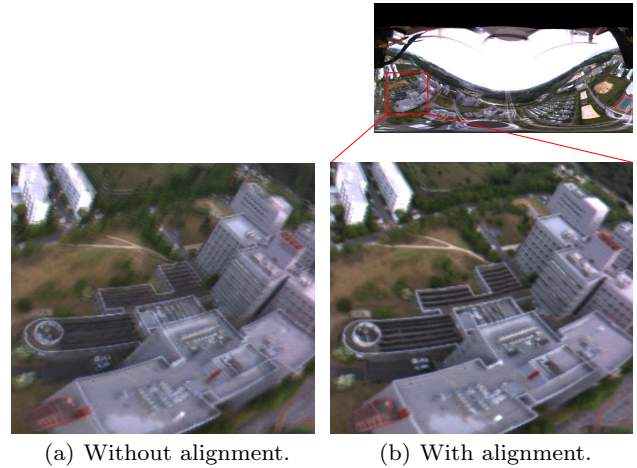


Fig. 5 HDR images with and without multi-exposure image alignment.

the number of pixels M_i corresponding to the exposure time s_{new_i} :

$$E_s = \sum_i (M_i \Delta_i). \quad (5)$$

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

The above process for determining 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.

3.3 HDR image generation from multi-exposure images

3.3.1 Alignment of multi-exposure images

Misalignments occur among multi-exposure images due to changes in the position and orientation of camera while capturing images at different exposure times. This causes blurring in the resulting HDR images, as illustrated in Fig. 5.

For an imaging aircraft moving at $5m/s$ and rotating $30^\circ/s$, the misalignment angles among multi-exposure images are calculated as shown in Table 2. In cases where the camera captures the scene from a great distance, the amount of misalignment is predominantly affected by changes in the orientation of the camera. We correct such misalignments by estimating the changes in camera orientation from the captured image sequence. Note that this correction method does not take position change into consideration, and will therefore ignore misalignments due to fast-moving objects

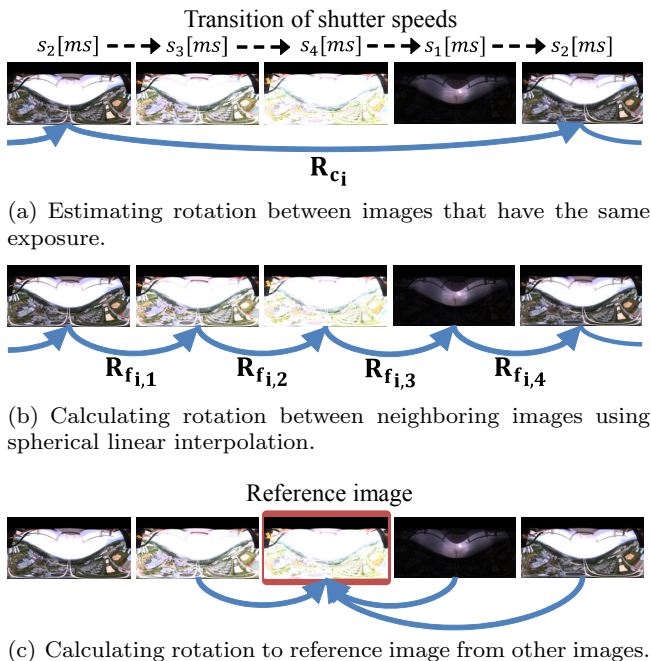


Fig. 6 Process for aligning multi-exposure images.

captured at short distances. Although our approach is quite simple compared with other alignment techniques that consider a position change (e.g., using a median threshold bitmap [Ward(2003)]), it is expected to be robust to misalignment due to orientation change. Fortunately, such existing multi-exposure alignment methods can simply be combined with our approach when significant camera motion occurs.

Fig. 6 depicts the steps of our multi-exposure images 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 Fig. 6(a). Among the multi-exposure images, those having the fewest saturated pixels (intensity of 255) and underexposed pixels (intensity of less than 16 in our implementation) are selected. The corresponding points between the selected images are determined by the Kanade-Lucas-Tomasi (KLT) tracker [Shi and Tomasi(1994)] and are

Table 2 Amount of misalignment due 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°

projected onto a unit sphere. \mathbf{p}_m and \mathbf{q}_m , respectively, denote the projected corresponding points. The parameters for rotation \mathbf{R}_{c_i} between the two images are estimated by nonlinearly minimizing the energy function E_l , defined as the sum of squared Euclidean distances between projected corresponding points $|\mathbf{p}_m, \mathbf{q}_m|$:

$$E_l = \sum_m |\mathbf{p}_m, \mathbf{q}_m|^2. \quad (6)$$

Note that RANSAC is used to reduce errors due to mismatches.

Next, as shown in Fig. 6(b), rotation $\mathbf{R}_{f_{(i,j)}}$ between every two adjacent images are calculated by interpolating \mathbf{R}_{c_i} using spherical linear interpolation.

Rotations to an arbitrary reference image from neighboring images are then calculated from $\mathbf{R}_{f_{(i,j)}}$, as shown in Fig. 6(c). Multi-exposure images are aligned to the reference image by transformation using the obtained rotation parameters. To generate aligned, multi-exposure images for an entire video sequence, $\mathbf{R}_{f_{(i,j)}}$ must first be calculated for the entire sequence, after which each frame in the sequence can be treated as the reference image.

3.3.2 HDR image generation

We use the HDR imaging method for multi-exposure images based on [Debevec and Malik(1997)], to compose the intensity of our LDR images. An ordinary method [Debevec and Malik(1997)] is modified to consider light attenuation due to the ND filters attached to the camera on top of the aircraft. If the response curve is linear, or linearized in advance, intensities in the HDR image I_h can be calculated using the LDR intensity I_l and the exposure time $t[s]$ of the LDR image:

$$I_h = \begin{cases} \gamma \frac{I_l}{t} & (\text{without ND filters}), \\ \gamma \frac{I_l}{\eta t} & (\text{with ND filters}). \end{cases} \quad (7)$$

The nominal light transmittance η of the ND filter is newly employed in addition to the existent HDR composition. Although the scale factor γ is meaningless when only relative pixel intensities are needed, it can be determined and used to calculate absolute radiance values [W/sr/m²]. The HDR intensities calculated from each set of multi-exposure images are composed in accordance with the method proposed by [Debevec and Malik(1997)].

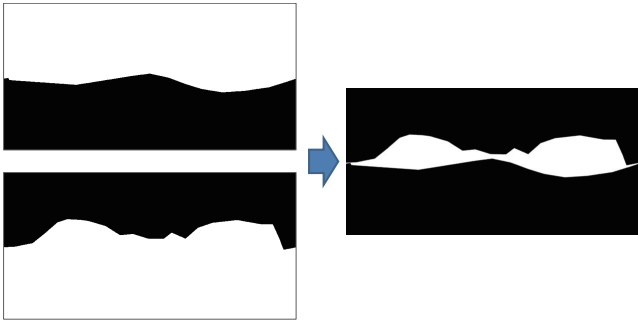


Fig. 7 Masks used for alignment of images from the two cameras. Top left: Black pixels indicate the missing area from the top of the aircraft. Bottom left: Black pixels indicate the missing area from the bottom of the aircraft. Right: White pixels indicate the area of overlap.

3.4 Composition of HDR images captured by two cameras

3.4.1 Alignment between the two cameras

The HDR images captured above and below 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 Fig. 7. Note that only one mask image is required for an entire sequence captured by a given camera, as the missing area for omnidirectional images does not change significantly. The area marked by white pixels on the right side of Fig. 7 indicates the overlap area where the scene has been captured by both cameras without occlusions. This area is defined as the conjunction of the negation of missing areas in the paired images.

Corresponding points are detected by applying the KLT tracker [Shi and Tomasi(1994)] to feature points in the overlap area of the image captured from the bottom of the aircraft. The parameters for rotation between the cameras from top to bottom are determined in the way described in Section 3.3.1.

3.4.2 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. The ND filters used in our experiment transmit less red light, as shown in Fig. 8. Such changes are corrected by estimating linear intensity transformation parameters from intensities in the overlap areas of images from the two cameras. The RGB values $(R_{top}(\mathbf{x}), G_{top}(\mathbf{x}), B_{top}(\mathbf{x}))$ for a pixel \mathbf{x} in the image captured from above the aircraft are converted to corrected pixel values $(R'_{top}(\mathbf{x}), G'_{top}(\mathbf{x}), B'_{top}(\mathbf{x}))$ us-



(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: $(R, G, B) = (95.1, 103.5, 111.9)$.

Fig. 8 Example of chromatic change due to ND filter. A pair of HDR images converted to the same exposure is shown in close-up, at region where the two images overlap.

ing the following linear transformations:

$$R'_{top}(\mathbf{x}) = \beta_r R_{top}(\mathbf{x}), \quad (8)$$

$$G'_{top}(\mathbf{x}) = \beta_g G_{top}(\mathbf{x}), \quad (9)$$

$$B'_{top}(\mathbf{x}) = \beta_b B_{top}(\mathbf{x}), \quad (10)$$

where β_r is estimated by

$$\beta_r = \frac{\sum_{\mathbf{x} \in A} \frac{R_{bot}(\mathbf{x})}{R_{top}(\mathbf{x})}}{N_A}, \quad (11)$$

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 overlap area, and N_A indicates the number of pixels belonging to A . β_g and β_b are estimated in a parallel manner.

3.4.3 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 overlap area of the full spherical image are determined by alpha blending the two images. From the intensities I_{top} and I_{bot} of a pixel in the overlap areas of the two images, the intensity I_{full} of the corresponding pixel of the full spherical image is calculated using

$$I_{full} = \alpha I_{bot} + (1 - \alpha) I_{top}, \quad (12)$$

where α varies linearly between zero on the upper boundary of the overlap area and unity on the lower boundary.

4 Results and discussions of generating full spherical HDR image and video

4.1 Generation of HDR image

To confirm that the HDR image generated using our proposed method reflects the real environment with rea-

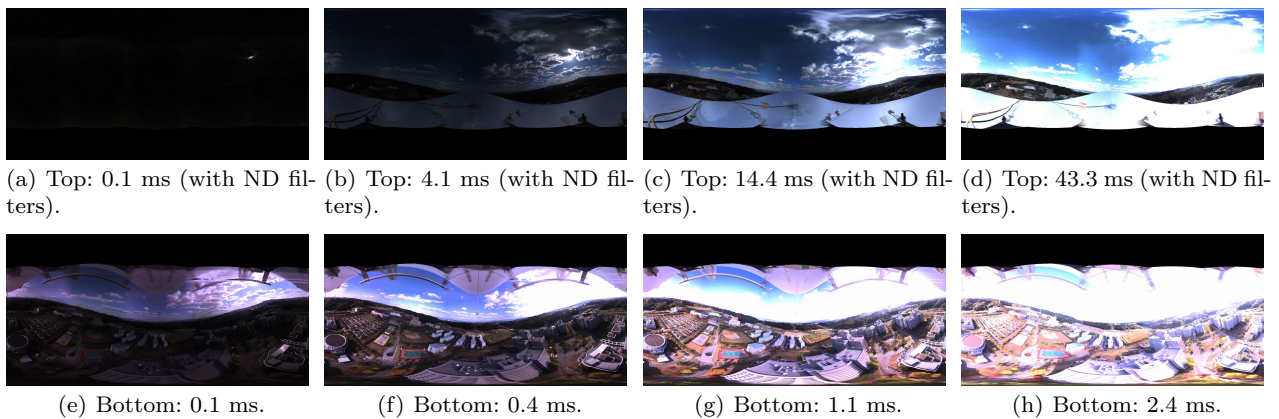


Fig. 9 Captured multi-exposure images and corresponding exposure time.

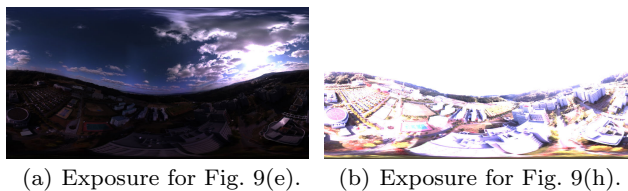


Fig. 10 Full spherical HDR images with cropped intensity and exposures equal to those of the multi-exposure images.



Fig. 11 Full spherical image tone-mapped in accordance with the method by [Reinhard et al(2002)].

sonable 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 a height of 130 m while capturing the multi-exposure images. $MAX_IN = 255$ and $\theta_{sh} = 192$ were used in this experimental environment.

The captured multi-exposure images and corresponding exposure times are shown in Fig. 9. Figs. 1(a) and 1(b) show HDR images from the top and bottom of the aircraft, respectively. The full spherical HDR image composed using Figs. 1(a) and 1(b) is shown in Fig. 1(c). Note that the two realigned and chromatically corrected images were successfully combined, leaving no

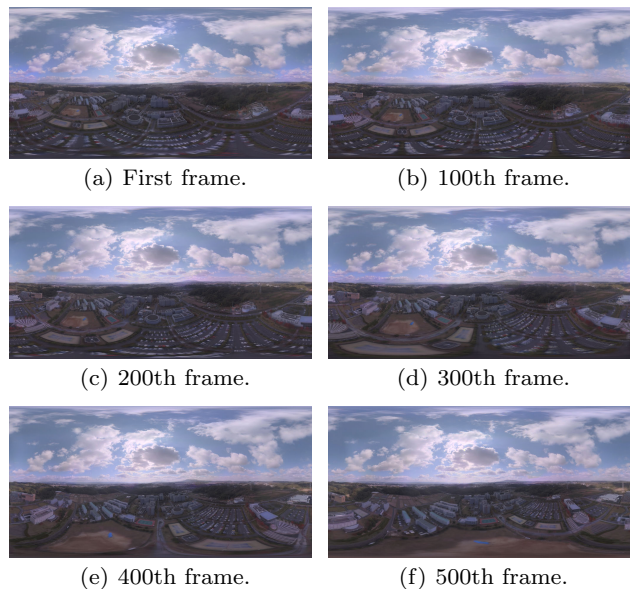


Fig. 12 Full spherical tone-mapped images generated from video frames. The images were aligned using estimated relative camera orientation, based on [Sato et al(2004)].

obvious artifacts. Fig. 10 shows the full spherical images for exposures equal to those in Figs. 9(e) and 9(h), respectively. From this, we can confirm that the full spherical HDR image satisfactorily combined the captured multi-exposure images. The full spherical HDR image can also be visualized by typical tone-mapping methods such as the one in [Reinhard et al(2002)], as shown in Fig. 11. Such tone-mapped images are suitable for immersive panoramas, which provide users a sense of looking around a location, and seeing textures of various radiances.

4.2 Generation of HDR video sequence

To generate full spherical HDR videos for flying-through applications, we applied our approach to a video sequence consisting of 500 frames. The aircraft was flown at a speed of 5–8 m/s and a height of 130 m while capturing the multi-exposure images. The frame rate of the generated HDR video was 16 fps, the same as that of the source multi-exposure video. Processing time was approximately 10 seconds per frame on a desktop PC with Core i7-2600 CPU (3.40 GHz, 4 Cores). The position and orientation of the camera beneath the aircraft were estimated using the structure-from-motion technique for omnidirectional cameras [Sato et al(2004)]. Fig. 12 shows samples of the full spherical frames. Note that the position and orientation of the video frames were successfully estimated. This information can be used for geometric registration between real and virtual objects in *augmented* immersive panoramas.

4.3 Evaluation of missing area completion using sky radiance model [Okura et al(2011)]

A promising method of spherical image completion is to fit a model of sky radiance to images acquired by a groundward camera and forgo use of a skyward camera altogether, as described in [Okura et al(2011)]. To demonstrate the effectiveness of this approach for IBL, we generated and completed a full spherical image using the All-Sky-Model [Igawa et al(2004)], which provides radiance parameters for common weather conditions. Note that the optimal parameters of the sky model were estimated from the intensity of the captured image using a method described in [Okura et al(2011)].

Fig. 13 shows the same frame as Fig. 12(b), but with the missing area completed using the sky model. Note that, from a subjective viewpoint, the estimated *smooth* textures in Fig. 13 may seem unnatural and decrease the immersive value of the image.

To investigate the effects when using full spherical images for IBL, we compared the estimated intensity shown in Fig. 13 with the acquired intensity. Note that 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 by Figs. 12(b) and 13. When Fig. 13 is used for IBL, virtual objects with Lambertian surfaces are rendered 0.218 times brighter than their captured intensity (Fig. 12(b)). Furthermore, in Fig. 12(b), the highest intensity (which affects the appearance of cast shadows) was ten times larger than in Fig. 13. These results indicate that the proposed approach, which employs an additional skyward camera



Fig. 13 Completion of missing area based on [Okura et al(2011)]

for capturing the actual intensity of the missing area, produces significantly better IBL rendering than current approaches using model-based sky area completion.

4.4 Discussion of auto-exposure control

We compared the auto-exposure control method in Section 3.2.2 with ordinary controls for HDR imaging. Most such controls, including that used by [Kang et al(2003)], first determine the base exposure automatically, and then calculate the remaining exposure values at fixed stops from the base exposure. So as not to saturate the scene, we used the shortest exposure time s_{c_1} [ms] as our base exposure. The other exposure times s_{c_2}, \dots, s_{c_4} were then calculated for every two stops, meaning that s_{c_i} was set $2^2 = 4$ 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 shows the calculated maximum intensity and quantization steps for each approach. The maximum intensity of each multi-exposure image is a relative value to the maximum intensity in the scene, defined as 1000. The quantization steps of the acquired multi-

Table 3 Quantization steps of the HDR image using the approach described in Section 3.2.2 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

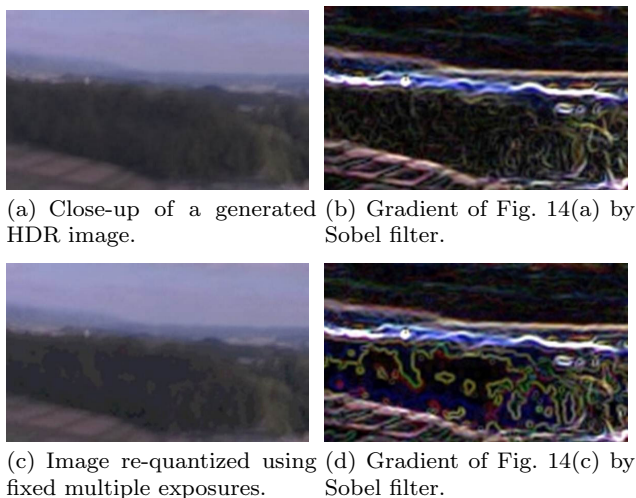


Fig. 14 Visualization of negative effect of large quantization steps.

exposure images (eight-bit) were converted to HDR intensity by dividing the maximum intensity by 255, in the same way as Δ in Section 3.2.2. A smaller value is better here, as it help prevent visual artifacts. A comparison of longer exposure times 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. Fig. 14 shows close-ups of a generated HDR image of the same frame shown in Fig. 11. Fig. 14(c) is the artificially re-quantized picture generated from this image using the exposure times of the conventional approach $s_{c_i} (1 \leq i \leq 4)$. Figs. 14(b) and 14(d) show the gradient images of Figs. 14(a) and 14(c), respectively. Note that banding appears in Fig. 14(d), mainly in low intensity areas. This indicates that our proposed approach can prevent visual artifacts produced using the conventional approach.

5 Displaying full spherical HDR image for immersive panoramas

In this section, we evaluate methods for displaying the full spherical HDR images generated in Section 3 based on subjective review of an immersive panorama application. In most uses of HDR images, they are converted using parameters determined from the image as a whole, such as maximum and logarithmic mean intensity. Immersive panorama applications, unlike ordinary applications, generally crop and convert a spherical image to a perspective image so that the entire image is not displayed at once. Thus, it is important to investi-

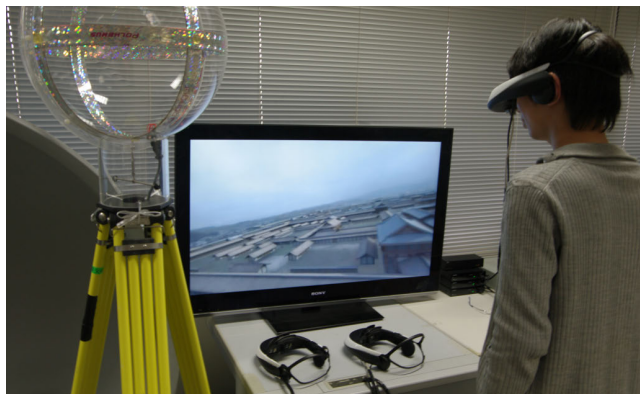


Fig. 15 Immersive system used in the experiments.

gate tone-mapping approaches suitable for such applications.

We conducted a subjective evaluation of four display methods using an immersive system comprising an head-mounted display (HMD) and a head tracking device [Onoe et al(1998)]. With this system, a spherical image is converted to a perspective image in real-time based on the user's view direction, as detected by an electro-magnetic sensor (Fastrak, Polhemus) mounted on the HMD (HMZ-T1, Sony Corp.). Fig. 15 depicts the system used.

The four display methods outlined below were evaluated Fig. 16 illustrates ideal tone curves for these methods.

1. LDR-like representation: This is a luminance transformation method that uses a linear tone curve, such as that shown in Fig. 16(a), to generate an LDR image with the same appearance as one generated by traditional LDR imaging.
2. Reinhard's tone-mapping [Reinhard et al(2002)]: This uses maximum and log-mean intensity to tone-map the HDR image. It is one of the most popular tone-mapping methods.
3. Region-wise tone-mapping: Aerial spherical images are divided into regions of bright sky and dark ground. Using the method of [Kitaura et al(2012)], these regions are then tone-mapping 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 by the immersive system based on the user's view direction. The intensity of a position in the scene varies according this view direction.

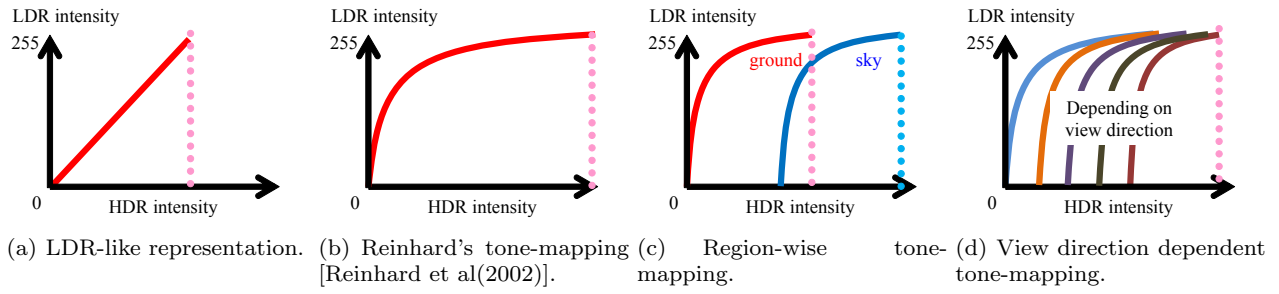


Fig. 16 Illustrative tone curves for four display methods.



Fig. 17 Image generated by LDR-like representation.



Fig. 18 Result using Reinhard's tone-mapping method [Reinhard et al(2002)].

5.1 Display methods for aerial spherical HDR images

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

5.1.1 LDR-like representation

Linear intensity transformation generates an image that looks the same as any generated by traditional LDR imaging. The tone curve is defined as

$$I_l = \begin{cases} 1.0 & (I_h > 1.0) \\ 2^v I_h & (\text{otherwise}), \end{cases} \quad (13)$$

where I_h is the intensity in the HDR image, I_l denotes the intensity of the LDR image normalized to the range $[0, 1]$, and v is the exposure value of the generated LDR image. We determined v so as to reduce overexposed and underexposed pixels in the generated LDR image. Underexposure was defined as $I_l < \frac{1}{16}$ and overexposure as $I_l = 1.0$ for this experiment. Note that, as seen in Fig. 17, many overexposed and underexposed pixels remained in the resulting image, which was converted from the HDR image captured in Section 4.

5.1.2 Reinhard's tone-mapping

The tone-mapping method proposed by [Reinhard et al(2002)] is composed of two steps.

In the first step, the intensity I_h of an HDR image is transformed to LDR intensity I_l by a function involving the log-mean \bar{I}_h :

$$I_s = \frac{\alpha}{\bar{I}_h} I_h, \quad (14)$$

$$I_l = \frac{I_s \left(1 + \frac{I_s}{I_{max}^2}\right)}{1 + I_s}. \quad (15)$$

Note that I_s is a scaled intensity using a scale factor α and I_{max} denotes the maximum intensity scaled by Eq. (14). The second step, so-called “dodging-and-burning” is applied to areas of low and high intensity to improve their visibility (detailed in [Reinhard et al(2002)]). Parameter α is automatically estimated by [Reinhard(2002)]. Fig. 18 shows the result of tone-mapping the same frame that used in Fig. 17.

5.1.3 Region-wise tone-mapping

Dividing images into sky and ground regions, we applied the approach proposed by [Kitaura et al(2012)] to tone-map these regions using separate parameters. To fully automate this process, we specialized it for aerial spherical imaging. Specifically, we used a priori knowledge of aerial images to support automatic Grab-Cut segmentation in place of user input of seed pixels. We assumed that the camera pose was estimated

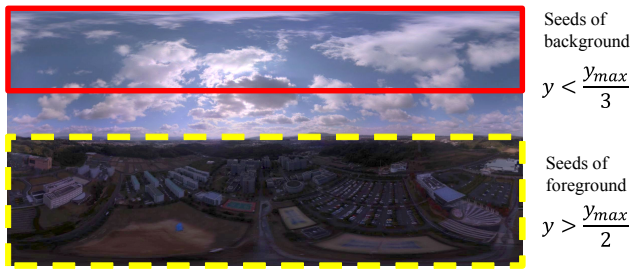


Fig. 19 Seed pixels for GrabCut [Rother et al(2004)]: y denotes the vertical component of the image coordinates; y_{max} is the height of the image.

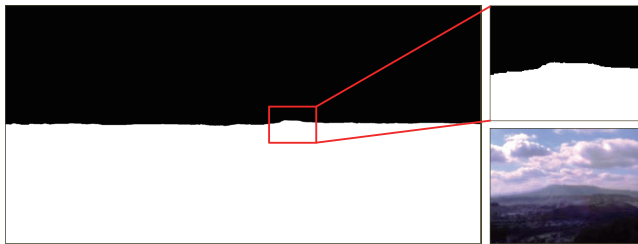


Fig. 20 Regions determined by GrabCut [Rother et al(2004)].



Fig. 21 Result of region-wise tone-mapping.

by structure-from-motion [Sato et al(2004)] in the same fashion as in Section 4.2, 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 Fig. 19. 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 successful separate of sky from ground, as shown in Fig. 20.

Region-wise tone-mapped images such as that shown in Fig. 21 are then produced by applying Reinhard's tone-mapping [Reinhard et al(2002)] to each region independently.



Fig. 22 Examples of view direction dependent tone-mapping: The intensity of the same position in the scene varies according to the view direction.

5.1.4 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 view direction. We applied a new approach similar to the tone-mapping for HDR video [Kang et al(2003)]. The approach consists of two processes:

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

Log-mean and maximum intensity were calculated from perspective images generated for each view direction (in this study, we calculated parameters for every degree of latitude and longitude). Eqs. (14) and (15) 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 et al(2002)]. Note that although this experiment used an offline process, it is possible to implement a fully real-time process by leveraging graphics processing unit (GPU)-based computation, as in other real-time tone-mapping proposals [Goodnight et al(2003)]. The results, seen in Fig. 22, show how the intensity at a given position in the scene can vary according to the user's view direction.

5.2 Subjective evaluation using the immersive system

To investigate the features of the above four display methods for aerial spherical HDR images, we conducted a subjective evaluation using an immersive system. Examinees used an HMD to view scenes generated by the 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, Vis)
 Q2: Did the tone, brightness, and appearance of textures appear natural? (Naturalness, Nat)

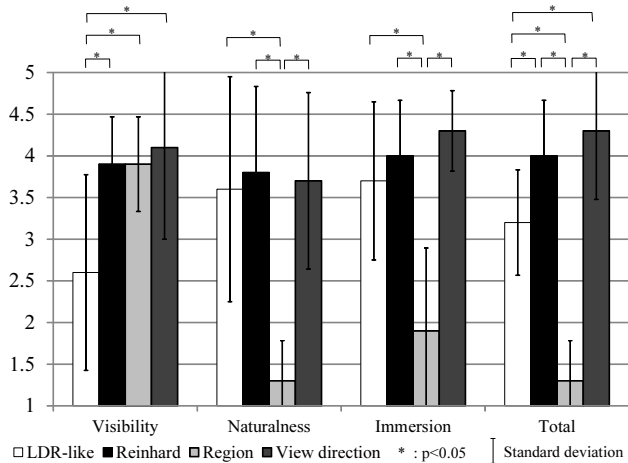


Fig. 23 Results of subjective experiment involving the four display methods by ten examinees. It also indicates the pairs that had significant differences ($p < 0.05$) calculated using the Tukey’s test, in which the type-I error rate and the degrees of freedom for each sample were set as 0.05 and $10 * 4 - 4 = 36$, respectively.

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

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

Note that the questions were originally given in Japanese. The examinees consisted of ten people in their twenties or thirties, amongst whom eight examinees were male and all of them had normal color vision. Although some of them were familiar with the concept of virtual reality and image processing, the aim of the experiment was kept secret during the experiment. No specific tasks were given to the examinees, except for watching the scenes. Further, the examinees had the option to change their view direction. The order in which the four display methods were presented was randomized, and examinees were allowed to switch and watch the display mode among the representations as many times as they wished, even while answering questions. The examinees were also allowed to comment freely after answering the questionnaire.

The results of the evaluation are shown in Fig. 23 (in detail, see Table 4). The figure also indicates the pairs that had significant difference ($p < 0.05$), 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 regards to naturalness.

5.3 Discussion of display methods

5.3.1 Effects of HDR imaging for immersive panorama

For Q4 in Section 5.2, 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 23). 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 [Moezzi(1997)], 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.

5.3.2 The best HDR image display method for immersive panorama

What is the best display method for HDR immersive panorama applications? 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.

There are no significant differences between Reinhard’s tone-mapping method and viewing-direction dependent tone-mapping. Free commentary from some examinees indicated that view direction-dependent tone-mapping improved visibility over Reinhard’s method, yet produced unnatural variations in intensity as view direction changed in real-time. Although both Reinhard’s tone-mapping method and view direction-dependent tone-mapping were thought to be promising representations for immersive display, the parameters related to human factors, such as acceptable rates of

Table 4 Specific results of the subjective experiment. The average (avg) and the standard deviation (stdev) of the score are shown in “avg(stdev)” format.

	LDR-like	Reinhard	Region	View direction
Vis	2.6(1.17)	3.9(0.57)	3.9(0.57)	4.1(1.10)
Nat	3.6(1.34)	3.8(1.03)	1.3(0.48)	3.7(1.06)
Imm	3.7(0.95)	4.0(0.67)	1.9(0.99)	4.3(0.48)
Tot	3.2(0.63)	4.0(0.67)	1.3(0.48)	4.3(0.82)

intensity variation, should be investigated to guide further development. If it can overcome this perceived deficiency, view-direction dependent tone-mapping may become a more popular display method.

6 Conclusion

This paper has focused on both the acquisition and display of aerial full spherical HDR images. The resulting HDR images can be used in immersive panoramic viewers, IBL, and other applications that use spherical images. The multi-exposure images used to generate these HDR images are captured by a pair of omnidirectional cameras mounted above and below an aircraft, and can even be used to generate full spherical HDR video, with position and orientation information available for superimposed virtual objects. We also investigated tone-mapping methods of full spherical HDR images using an immersive panorama system using an HMD. It was confirmed that the HDR imaging is well suited to such immersive applications, and the results also show that a tone-mapping method that depends on the user's viewing direction as well as an ordinary tone-mapping yields good results with respect to critical human factors.

7 Future Work

Although we have developed a complete system for acquiring aerial full spherical HDR images, there are some issues that should be investigated/developed further.

Our system employs a new exposure control approach for multi-exposure imaging, which selects exposures such that the quantization steps become small. This approach can be applied for general multi-exposure imaging in principle. To utilize our approach for general situations, the performance should be compared with more sophisticated exposure control approaches, in addition to a simple method using fixed-multiple exposures as in Section 4.4.

With the recent growth of hardware technology, some CCD/CMOS sensors can directly acquire HDR (e.g., 12- or 15-bit) scenes. However, it is not enough to capture a whole dynamic range of outdoor scenes, as well as a whole direction of the scenes in a single shot, in which the dynamic range only of the sky can be up to approximately $1 : 2^{17}$ (may become larger for a whole scene including the ground). However, such HDR imaging sensors would be a good substitute for using multi-exposure imaging in the near future.

It is a good future direction to investigate/develop more state-of-the-art display methods for HDR immersive panorama. The experiment, which was a compari-

son among some relatively old tone-mapping methods, showed that there remains room for improvement with regard to the view-direction dependent approach. Uses for web-browser-based immersive panorama systems, similar to Google Street View, can also be taken into account in addition to using an HMD.

We also intend to use full spherical HDR images for photorealistic superimposition of virtual objects, based on camera position and orientation information. Such applications are promising for mixed-reality-based visualization such as cultural heritages.

Acknowledgements This research was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (A), No. 23240024, Grant-in-Aid for JSPS Fellows No. 25-7448, and by the "Ambient Intelligence" project funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT).

References

- [Anguelov et al(2010)] Anguelov D, Dulong C, Filip D, Frueh C, Lafon S, Lyon R, Ogale A, Vincent L, Weaver J (2010) Google street view: Capturing the world at street level. *IEEE Comput Mag* 43(6):32–38
- [Chen(1995)] Chen SE (1995) Quicktime VR: An image-based approach to virtual environment navigation. In: *Proc ACM SIGGRAPH'95*, Los Angeles, CA, pp 29–38
- [Debevec(1998)] Debevec P (1998) Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: *Proc ACM SIGGRAPH'98*, Orlando, FL, pp 189–198
- [Debevec et al(1996)] Debevec P, Taylor C, Malik J (1996) Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In: *Proc ACM SIGGRAPH'96*, New Orleans, LA, pp 11–20
- [Debevec and Malik(1997)] Debevec PE, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: *Proc ACM SIGGRAPH'97*, Los Angeles, CA, pp 369–378
- [Goodnight et al(2003)] Goodnight N, Wang R, Woolley C, Humphreys G (2003) Interactive time-dependent tone mapping using programmable graphics hardware. In: *Proc 14th Eurographics Workshop on Rendering*, Leuven, Belgium, pp 26–37
- [Goodrich et al(2008)] Goodrich MA, Morse BS, Gerhardt D, Cooper JL, Quigley M, Adams JA, Humphrey C (2008) Supporting wilderness search and rescue using a camera-equipped mini UAV. *J Field Robotics*, 25(1):89–110
- [Grossberg and Nayar(2003)] Grossberg MD, Nayar SK (2003) High dynamic range from multiple images: Which exposures to combine? In: *Proc IEEE Workshop on Color and Photometric Methods in Comput Vis (CPMCV)*, Nice, France, pp 1–8
- [Hasinoff et al(2010)] Hasinoff SW, Durand F, Freeman WT (2010) Noise-optimal capture for high dynamic range photography. In: *Proc 23rd IEEE Conf Comput Vis Pattern Recognit (CVPR'10)*, San Francisco, CA, pp 553–560

- [Herwitz et al(2004)] Herwitz S, Johnson L, Dunagan S, Higgins R, Sullivan D, Zheng J, Lobitz B, Leung J, Gallmeyer B, Aoyagi M, Slye R, Brass J (2004) Imaging from an unmanned aerial vehicle: Agricultural surveillance and decision support. *Comput Electronics in Agriculture*, 44:49–61
- [Igawa et al(2004)] Igawa N, Koga Y, Matsuzawa T, Nakamura H (2004) Models of sky radiance distribution and sky luminance distribution. *Sol Energy*, 77(2):137–157
- [Ikeda et al(2003)] Ikeda S, Sato T, Yokoya N (2003) High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system. In: *Proc. 2003 IEEE Int Conf Multisens Fusion Integration Intelligent Syst (MFI'03)*, Tokyo, Japan, pp 155–160
- [Kang et al(2003)] Kang SB, Uyttendaele M, Winder S, Szeliski R (2003) High dynamic range video. *ACM Trans Graph (Proc. ACM SIGGRAPH'03)*, 22(3):319–325
- [Kawai et al(2009)] Kawai N, Sato T, Yokoya N (2009) Image inpainting considering brightness change and spatial locality of textures and its evaluation. In: *Proc Third Pacific-Rim Symp Image and Video Technology (PSIVT'09)*, Tokyo, Japan, pp 271–282
- [Kawai et al(2010)] Kawai N, Machikita K, Sato T, Yokoya N (2010) Video completion for generating omnidirectional video without invisible areas. *IPSJ Trans Comput Vis Appl* 2:200–213
- [Kitaura et al(2012)] Kitaura M, Okura F, Kanbara M, Yokoya N (2012) Tone mapping for HDR images with dimidiated luminance and spatial distributions of bright and dark regions. In: *Proc SPIE, Burlingame, CA*, vol 8292, ArticleNo 829205
- [Kopf et al(2010)] Kopf J, Chen B, Szeliski R, Cohen M (2010) Street slide: Browsing street level imagery. In: *ACM Trans Graph (Proc. ACM SIGGRAPH'10)*, 29(4), ArticleNo 96
- [Moezzi(1997)] Moezzi S (1997) Special issue on immersive telepresence. *IEEE Multimedia* 4(1):17–56
- [Okura et al(2011)] Okura F, Kanbara M, Yokoya N (2011) Fly-through Heijo palace site: Augmented telepresence using aerial omnidirectional videos. In: *Proc ACM SIGGRAPH'11 Posters*, Vancouver, BC, ArticleNo 78
- [Onoe et al(1998)] Onoe Y, Yamazawa K, Takemura H, Yokoya N (1998) Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Comput Vis Image Underst* 71(2):154–165
- [Reinhard(2002)] Reinhard E (2002) Parameter estimation for photographic tone reproduction. *J Graph Tools* 7(1):45–51
- [Reinhard et al(2002)] Reinhard E, Stark M, Shirley P, Ferwerda J (2002) Photographic tone reproduction for digital images. *ACM Trans Graph (Proc. ACM SIGGRAPH'02)*, 21(3):267–276
- [Reinhard et al(2010)] Reinhard E, Heidrich W, Debevec P, Pattanaik S, Ward G, Myszkowski K (2010) High dynamic range imaging: Acquisition, display, and image-based lighting. Morgan Kaufmann, San Francisco, CA
- [Rother et al(2004)] Rother C, Kolmogorov V, Blake A (2004) Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans Graph (Proc. ACM SIGGRAPH'04)*, 23(3):309–314
- [Sato et al(2004)] Sato T, Ikeda S, Yokoya N (2004) Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. In: *Proc Eighth European Conf Comput Vis (ECCV'04)*, Prague, Czech Republic, 2:326–340
- [Shi and Tomasi(1994)] Shi J, Tomasi C (1994) Good features to track. In: *Proc. 1994 IEEE Conf Comput Vis Pattern Recognit (CVPR'94)*, Seattle, WA, pp 593–600
- [Shum et al(2003)] Shum H, Kang S, Chan S (2003) Survey of image-based representations and compression techniques. *IEEE Trans Circuits and Syst for Video Technology* 13(11):1020–1037
- [Stumpfel et al(2004)] Stumpfel J, Tchou C, Jones A, Hawkins T, Wenger A, Debevec P (2004) Direct HDR capture of the sun and sky. In: *Proc Third Int Conf Comput Graph, Virtual Reality, Visualization and Interaction in Africa (AFRIGRAPH'04)*, Stellenbosch, South Africa, pp 145–149
- [Ward(2003)] Ward G (2003) Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *J Graph Tools* 8(2):17–30
- [Zomet et al(2003)] Zomet A, Feldman D, Peleg S, Weinshall D (2003) Mosaicing new views: The crossed-slits projection. *IEEE Trans Pattern Anal Machine Intell* 25(6):741–754