Denoising and Texturing for Three Dimensional Modeling Utilizing Laser Attribute

大石, 修士

https://doi.org/10.15017/1500752

出版情報：九州大学, 2014, 博士（工学）, 課程博士
バージョン：

Denoising and Texturing for Three Dimensional Modeling Utilizing Laser Attribute

Shuji Oishi

Graduate School of Information Science and Electrical Engineering,
Kyushu University

Presented for the degree of Doctor of Philosophy
and the Diploma of Kyushu University.

January 2015

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Ryo Kurazume. His immense knowledge and expertise helped me throughout this work while his kindness and patience allowed me to grow as a research scientist. I also appreciated a great freedom to plan and execute my ideas, and that is reflected in my unique research topics. I believe that he has been the best supervisor for a student would wish for, and it was an honor to be one of those who had an opportunity to work with him.

I am deeply grateful to Prof. Tsutomu Hasegawa, Prof. Kazunori Umeda, Assoc. Prof. Yumi Iwashita, Assoc. Prof. Ken'ichi Morooka, Assoc. Prof. Hajime Nagahara, and Assist. Prof. Tokuo Tsuji for their guidance, encouragement, and insightful comments.

I also would like to thank my fellow Labmates in Intelligent Robots and Vision System (I.R.V.S.) in Kyushu University, especially, Koji Uchino, Masahide Tanaka, Soichiro Uto, Naoaki Kondo, Yoonseok Pyo, and Hojung Jung for valuable discussions and suggestions.

Last, but not least, I would like to dedicate this thesis to my family. My sincere thanks also goes to my wife Sayoko for her encouragement, support, and patience.

Abstract

3D geometry measurement in the real world is one of the most significant tasks in research fields such as robotics, computer graphics, and computer vision. Time-of-flight ranging is one of the most efficient range measurement technologies which can provide accurate 3D geometric information stably without being affected by lighting conditions. Time-of-flight range sensors basically emit light/laser toward a target object, and calculate a distance between the sensor and the target based on a round-trip time of the emitted light/laser. Due to the measurement principle, time-of-flight based range measurement has several important advantages. Its collinearity of illumination and observation leads avoiding shadow effects and providing accurate contour in the range image. In addition, complex image processing, such as extracting visual features and finding matching points among input images on which stereo triangulation relies, is not required to obtain range data. Moreover, since time-of-flight sensors achieve the accurate ranging by actively emitting light/laser to the scene, they are less subject to lighting conditions which means they can perform even in nighttimes. These characteristics enable time-of-flight sensors to work stably even in outdoor environments, and they have been widely used for various applications such as remote control of a rescue robot in a hazardous environment or 3D modeling for digital archives of cultural heritages.

For further accurate modeling and understanding of the 3D world, this dissertation addresses on enhancing the functionality of 3D modeling with **a time-of-flight laser scanner utilizing attributes of the laser reflectivity**. A noteworthy characteristic of laser scanner is that they obtains a power of the reflected laser (reflectivity) as a side-product of the range data. Laser reflectivity indicates an intensity on the surface of target objects under a single-frequency light source, and a reflectance image, which is a collection of laser reflectance depicted as a grayscale image, contains rich appearance information about the target object. Therefore, laser scanners perform range and intensity measurements for each point simultaneously, and geometric and appearance information principally aligned together is available requiring no calibration setting which is generally essential for a range and image sensor combination.

We take advantage of the benefits of the laser reflectivity and develop novel techniques that enhance the functionality of laser scanners. This dissertation actually addressed four different issues: Range image smoothing using trilateral filter, Range image completion with belief propagation, Manual/Automatic colorization for 3D geometric models, and Texture synthesis for hole-free texture mapping. A variety of simulations and experiments demonstrated the validity of the proposed algorithms.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Human vision versus robot vision

Five senses - Sight, Touch, Taste, Smell, and Hearing - play essential roles in daily life allowing us to monitor our bodies and interact with the external environment. Compared with the other senses, we significantly rely on visual perception. Visual perception provides us with a wealth of optical information of objects in the world around us, such as their material properties and three dimensional (3D) shapes. This enables us to safely perform interaction with surrounding environments such as activities of daily living (ADL) including walking around, handling objects, driving a car and so on.

While these activities are easy tasks for us, it is quite difficult for robots to mimic human behaviors observing surrounding environments with vision sensors. Here, we have a rethink on ”what humans and robots see” in order to discuss the difference between our eyes and robot vision.

Human eye plays an important role as our window to the world allowing us to see objects. The most important components of human eye would be a lens and a retina: The lens concentrates light rays reflected off an object onto the retina, and the retina transforms the light into neurological signals that a brain can interprets. Finally, the neurological signals are formed as an image of the object by the brain. That is a simple human vision mechanisms that take place in a human eye.

In order to create an artificial eye, an enormous amount of research has been done, and a variety of vision sensors has been available recently. The most basic and standard examples of the vision sensors would be a digital camera. A digital camera typically consists of a lens and a CCD/CMOS image sensor, and it captures a scene by concentrating incident light onto the CCD/CMOS image sensor with the lens. The image sensor has an array of millions of tiny photosites, and is able to convert the incident light hitting the photosites array into intensity signals. This mechanism is analogous to that of human vision. Although the resolution of a digital camera generally has limitations depending on the physical structure of the image sensor, the resolution of state-of-the-art digital cameras is reaching hundreds of mega pixel and achieving sufficient quality compared with human eye’s resolution [5].

Now, we have well-developed artificial vision sensors. So the question is ”whether robot vision that has the same ability to our eyes has been developed using these vision sensors? - Not yet.”

Take a look around the room that you’re currently in. You may see your phone, personal computer, bag, or so on, and at the same time, you may notice you naturally recognize these objects and imagine their geometric structures. Our vision system enables us to perform a wide variety of tasks that need object recognition and handling, and the human vision is based on highly sophisticated 3D perception. Humans are able to see the spatial relationship among objects in a scene capturing different views from each eye and combining the information based on binocular stereo principal, and this enables us to perceive the 3D world. On the other hand, a digital camera allows us to capture a high-resolution two-dimensional (2D) image of a scene, however, in order to perform tasks such as grasping objects and walking without collision, raw information derived from a 2D image is not sufficient.

On the other hand, a digital camera allows us to capture a high-resolution two-dimensional (2D) image of a scene, however, in order to perform tasks such as grasping objects and walking without collision, raw information derived from camera images is not sufficient. As demonstrated in *trompe l'oeil* art works, recovering 3D shape from a single 2D image is generally an ill-posed problem. In other words, an object has different appearances from different viewpoints due to its 3D geometry, and 3D perception is essential for understanding a surrounding environment correctly.

Measuring 3D geometric information of a scene is one of the most essential issue in many research fields. During last decade, a variety of 3D measurement sensors has been developed and today leaving the laboratory to perform everyday tasks, at the same time, a lot of researches have been done on 3D modeling and reconstruction. An introduction to 3D measurement is given in the following section including reviews of major 3D measurement technologies.

1.2 3D geometry measurement

3D geometry measurement in the real world is one of the most significant tasks in research fields such as robotics, computer graphics, and computer vision. For example, acquiring accurate 3D geometric maps in unknown scenes is indispensable for many robotics applications including mobile robot localization and navigation or object recognition and manipulation. Augmented reality (AR), which enriches the real world overlaying 3D models and digital information on video sequences, as well needs

Figure 1.1: 3D reconstruction results in ”Building Rome in a Day”[1]

3D geometric information in a scene so that it could appropriately represent interactions such as collision and occlusion among the virtual 3D models and real objects in the scene. Precise and dense 3D geometric measurement itself is also an essential technology to create virtual 3D models of cultural heritages such as famous buildings and art works for digital archive and museum. Known examples of such application include ”Building Rome in a Day”[1]. In this project, Agarwal et al. worked on a problem of reconstructing the entire Rome city based on the enormous amount of images collected from Flickr.com. They developed a parallel distributed system that was able to find corresponding points among these massive images quickly and to compute 3D structure (point clouds) of the city using the bundle adjustment in a day (Fig.1.1).

As much attention has been attracted to above-mentioned applications, range measurement technology has been developed rapidly throughout the years, and a variety of techniques/devices has been developed. Basically range sensing denotes a technique that measures the distance from the sensor position toward a certain point sampled from an object’s surface in a scene. According to the type of device/methodology, range sensing techniques fall into several different categories. This chapter discusses the characteristics of some 3D measurement techniques and explore the literature of developments throughout the years in some depth.

1.2.1 Stereo triangulation

An digital camera captures a 2D image of a scene where 3D world is projected onto the 2D image plane through the aperture of the camera. In order to determine a position of a certain point in 3D space, stereo triangulation solves an inverse problem of this 3D-2D projection based on two or more images like our vision system. Given a pair of images, binocular stereo provides a fundamental solution to this problem in the simplest geometrical case where the images are on the same plane. In addition, when the relative pose and position between these images are not available, it is necessary to determine the extrinsic parameters simultaneously. This problem is called ”Simultaneous Localization and Mapping (SLAM)” in robotics[6], ”Structure from Motion (SfM)” in computer vision[7], and a number of studies has addressed 3D reconstruction exploiting these techniques. Here, the triangulation principle underlying stereo vision is introduced, and we also discuss characteristics of passive/active stereo 3D reconstruction approaches.

Passive stereo with two/multiple camera images

Binocular stereo is a typical passive range sensing technique that simulates the human visual system for 3D perception. Every visible 3D point of the scene surface is restricted to a ray which passes through the camera center and its projection on the 2D image. Given a pair of images taken from different viewpoints, binocular stereo detects the 3D location of each visible point by determining an intersection of two rays from the camera centers based on triangulation. Assuming a position of a 3D point is estimated correctly, it should have the same color for its 2D projections to stereo images. Therefore, the main problem to be solved in binocular stereo is finding corresponding pixel pairs in two stereo images based on visual cues. Since triangulation crucially depends on the corresponding pairs, ambiguous correspondences may lead to inconsistent interpretation of the target scene and result in poor 3D reconstruction quality. Much attention has been attracted to the ”stereo correspondence problem”,

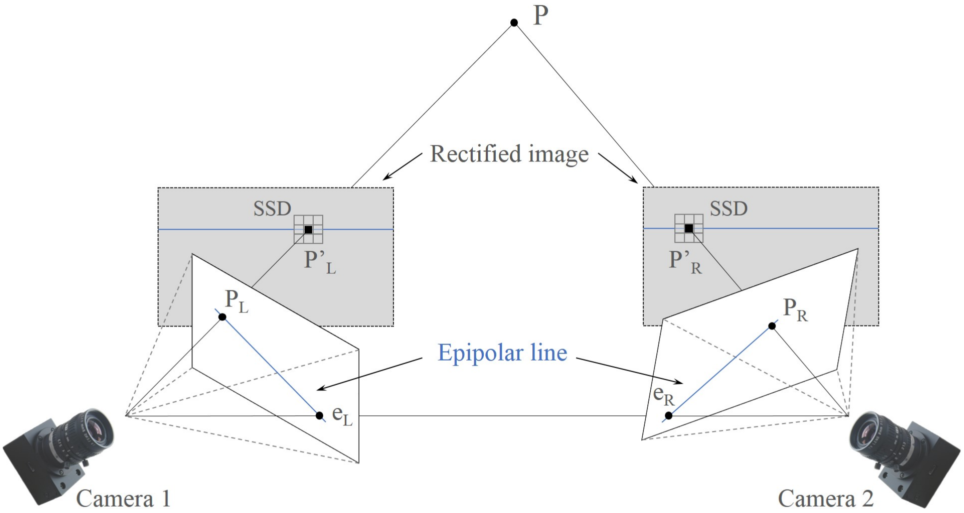


Figure 1.2: Binocular stereo principle

and an enormous work has been done to achieve high-accuracy 3D measurement over recent decades.

As explained above, since a position of a visible 3D point is located on a ray which passes a viewpoint and its projection onto the 2D image plane, the stereo correspondence search is also restricted to an epipolar line which denotes the ray projected onto the second view. Stereo correspondences are established by finding the most similar pixel- or area-pairs along the epioplar lines based on visual cues, and a variety of descriptor can be used for it[8][9][10]. The most typical method to measure similarity for a pixel pair would be the sum of squared differences (SSD). This method evaluates the similarity of a pixel pair in stereo images by calculating color differences around the pixels, and detects a correspondence along a epipolar line which minimizes SSD(Fig.1.2).

On the other hand, since binocular stereo basically assumes that every scene element is visible and appears to be similar to both cameras, scene elements which are visible to only one of them or whose appearances look different according to viewpoint pose a challenge to the stereo matching problem. For the difficulty of binocular stereo, the use of multiple camera images gives a solution. The original idea was introduced as ”multiple-baseline stereo” that improves the stereo matching reliability by integrating a correspondence search result of each binocular stereo pair[11]. Binocular stereo generally determines a corresponding pair for each pixel calculating the similarity scores along a epipolar line, however, the corresponding pixel cannot be determined uniquely when several pixels have the same minimum score. To deal with this problem, Okutomi et. al.[11] sums up SSDs of binocular stereo pairs with different baselines. This approach allows us to reduce the stereo matching ambiguity and avoid mismatches by finding a minimum of the sum of SSDs (SSSD) which is supposed to be a global minimum. Defining a SSSD and determining its minimum for each pixel in a reference image, it provides a high-quality depth map.

Approaches mentioned earlier addressed only stereo correspondence problem for the simplest geometrical case where stereo image planes are parallel or the case input images are rectified beforehand using the extrinsic parameters[12]. When we actually estimate 3D shape of a scene using multiple images taken from different viewpoints or an image sequence taken from a moving camera, calibration or camera tracking must be performed as well in order to obtain the camera position and pose for each image.

The approach that reconstructs a 3D scene and estimates viewpoints simultaneously from an image sequence is known as ”Structure from Motion (SfM)” and ”Simultaneous Localization and Mapping (SLAM)”. The key idea is as follows: Given a set of images of a target scene, reconstruct the 3D geometry along with camera parameters according to correspondences among the multi-viewpoint images. It is generally decomposed into two tasks: Correspondence determination and 3D structure/Camera parameter estimation. In the former step, point-to-point correspondences among a set of images taken from multiple viewpoints are determined based on photometry constraints. Robust descriptors such as SIFT[8] and SURF[9] are often used to describe each salient point in the images, and the point-to-point correspondences are build based on feature matching. The latter step estimates the scene geometry and camera motions according to the correspondences using Singular Value Decomposition (SVD), Direct

Linear Transformation (DLT), or bundle adjustment. In recent approaches, bundle adjustment is mainly used for the latter step due to its capability of accurate parameter estimation[13].

Active stereo with structured light

Passive stereo is capable of measuring a dense depth map when the target object is well-textured, however, it is difficult to determine corresponding stereo pairs for objects with low-texture due to the visual ambiguity. In addition, it is subject to lighting conditions and that sometimes makes difficult to find correspondences in stereo images. In contrast, active stereo allows us to avoid the ambiguous correspondences by replacing one of the cameras with a controlled illumination device. It basically projects structured light patterns onto the scene in order to generate virtual textures, and measures depth values based on the reflection captured with a digital camera via the same principle as passive binocular stereo.

So far, a variety of illumination patterns has been proposed such as line/stripe[14], grid[15], sinusoid[16], and random dots[17]. For example, a pair of a slit laser and a rotating mirror can be used for sequential scanning of an object's surface. The slit laser projected in a certain direction strikes the surface and is observed in a camera image. According to the stereo principle, bright pixels indicate intersections of the projection rays and camera rays, and disparities can be easily calculated by finding corresponding stereo pairs from the bright pixels in a camera image.

While the active stereo provides accurate 3D models avoiding the ambiguous correspondences, it has several drawbacks: First, it is difficult to find corresponding pairs for specular or black surfaces where the projected illumination does not reflect in the camera direction and cannot be observed in a camera image. In addition, the active stereo needs more time compared to the passive binocular stereo for projecting illumination patterns.

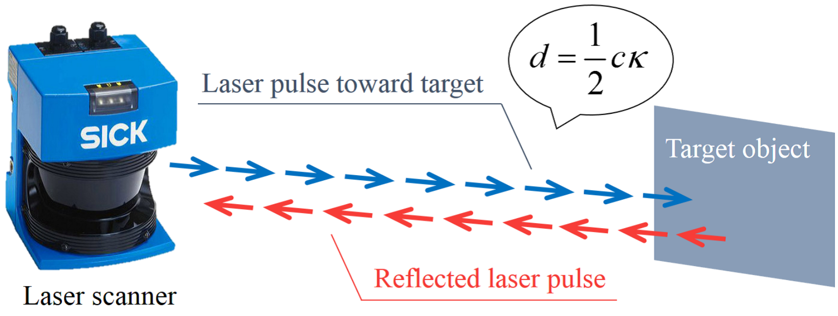


Figure 1.3: Time-of-flight pulse propagation time measurement principle

1.2.2 Time-of-flight ranging

Time-of-flight(ToF) range measurement technology is used in active optical sensors such as laser scanners and ToF cameras. The basic principle can be summarized as follows: First, light/laser is emitted from an optical unit toward a target object, and returns to a receivers in the time-of-flight sensor after traveling to the target's surface and reflecting on it. Second, a distance between the sensor and the target is calculated based on a time difference between light/laser emission time and reception time.

In general, there are three time-of-flight principles: Pulse propagation time measurement, Continuous wave modulation, and Range gated imaging.

Pulse propagation time measurement

This ranging method measures the time of flight directly: Emitting short-pulsed light toward a target object and estimating the distance d to the object based on the velocity of light c and the propagation time κ from the light emission to reception as follows (Fig.1.3).

$$d = \frac{1}{2}c\kappa \tag{1.1}$$

Continuous wave modulation

A depth value is obtained by calculating the phase shift based on the cross-correlation $C(\tau)$ of emitted wave signal $s(t)$, whose period is T, and reflected signal $r(t)$ by the scene. Exploiting a sinusoidal signal as incident signal $s(t)$, reflected signal $r(t)$ and the cross-correlation $C(\tau)$ is calculated as follows[2]:

$$\begin{aligned} s(t) &= \cos(\omega t) && (1.2)\\ r(t) &= \alpha\cos(\omega t+\phi)+\beta && (1.3)\\ C(\tau) &= r(t)\otimes s(t) = \lim_{T\to\infty}\int_{-T/2}^{T/2} r(t)s(t+\tau)dt && (1.4)\\ &= \frac{\alpha}{2}\cos(\omega\tau+\phi)+\beta && (1.5) \end{aligned}$$

where ω denotes the modulation frequency, α is an amplitude of the reflected signal, β is an offset due to the ambient illumination, and ϕ is the phase offset relating to the distance.

It can be demodulated utilizing the correlation $C(\tau)$ sampled with four different phase offsets ($\tau_i = i\cdot\frac{2}{\pi}, i = 0,\ldots,3$). Based on these four correlation values, the amplitude α, offset β, and phase offset ϕ can be obtained as follows (Fig.1.4):

$$\begin{aligned} \alpha &= \frac{1}{2}\sqrt{(C(\tau_3)-C(\tau_1))^2+(C(\tau_0)-C(\tau_2))^2} && (1.6)\\ \beta &= \frac{1}{4}\sum_{i=0}^{3}C(\tau_i) && (1.7)\\ \phi &= \arctan\left(\frac{C(\tau_3)-C(\tau_1)}{C(\tau_0)-C(\tau_2)}\right) && (1.8) \end{aligned}$$

Finally, the distance d is calculated according to the phase offset ϕ.

$$d = \frac{1}{4\pi\omega}c\phi \qquad (1.9)$$

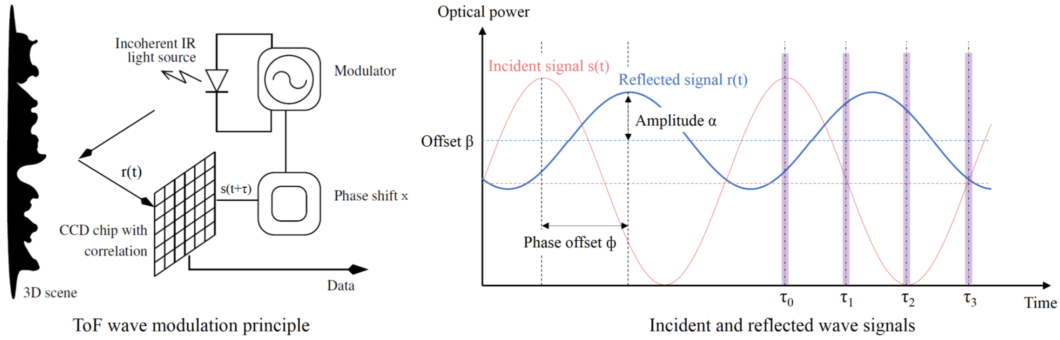


Figure 1.4: Time-of-flight continuous wave modulation principle [2]

Range gated imaging

Fast gating with a shutter in front of the image sensor enables us to measure 3D information. The concept of this technique is based on projecting ”light wall” which has a certain width and moves along the field of view. When the light wall is reflected by a scene, it gets distorted according to the target shape and returns to the sensor carrying an imprint of the target.

The 3D structure can be extracted from the deformed light wall using a built-in shutter in front of the image sensor that blocks incoming light. The shutter opens for the same period $T_{pulse/gate}$ as the light wall emission, and each receptor on the image sensor may collect the incoming light with a short delay T_{delay} during the gate time $T_{pulse/gate}$. The pixel intensity $I_{collected}$, which denotes the amount of collected light by a receptor, is in inverse proportion to the distance. The distance d is calculated according to the ratio of $I_{collected}$ and $I_{blocked}$, which denotes the amount of light blocked by the shutter, as follows (Fig.1.5):

$$d = \frac{1}{2} c T_{pulse/gate} \frac{I_{blocked}}{I_{collected} + I_{blocked}} \tag{1.10}$$

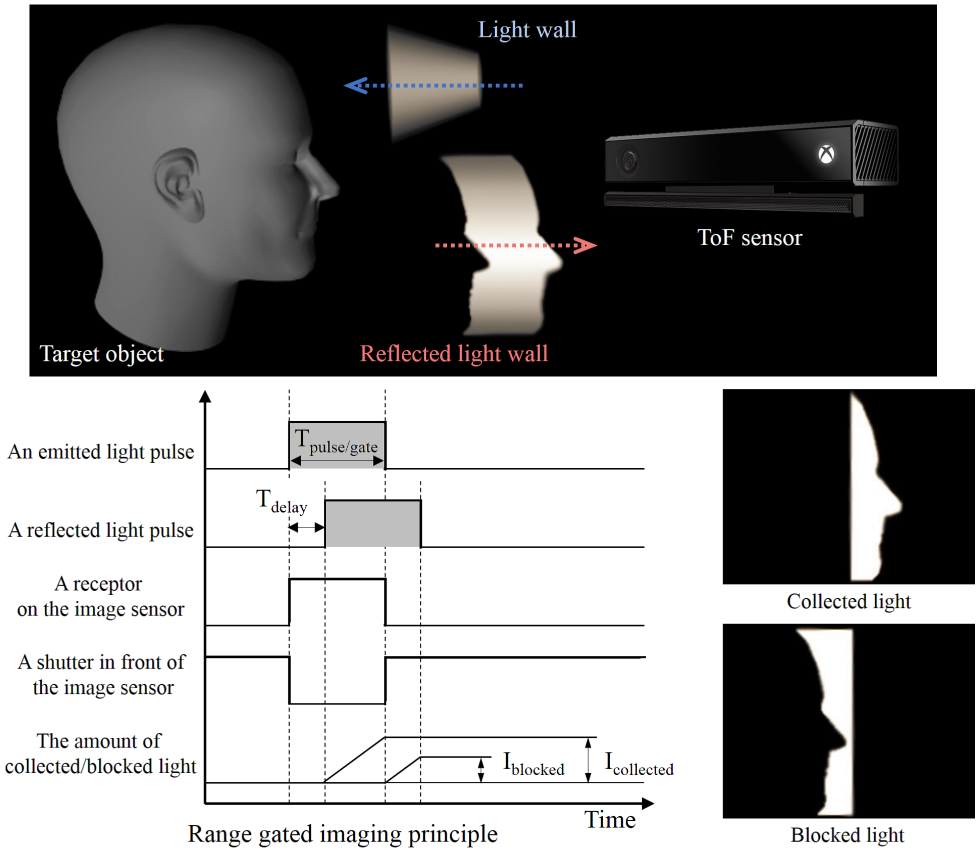


Figure 1.5: Time-of-flight range gated imaging principle [3]

1.3 Time-of-flight laser scanner and laser reflectivity

Due to the measurement principle, time-of-flight based range measurement has several important advantages. Its collinearity of illumination and observation leads avoiding shadow effects and providing accurate contour in the range image. In addition, complex image processing, such as extracting visual features and finding matching points among input images on which stereo triangulation relies, is not required to obtain range data. Moreover, since time-of-flight sensors achieve the accurate ranging by actively emitting light/laser to the scene, they are less subject to lighting conditions which means they can perform even in nighttimes. These characteristics enable time-of-flight sensors, especially **laser scanners**, to work stably even in outdoor environments, and they have been widely used for various applications such as remote control of a rescue robot in a hazardous environment[18] or 3D modeling for digital archives of cultural heritages[4].

As explained above, **a laser scanner**, one of time-of-flight sensors based on pulse propagation time measurement, obtains range data by emitting a laser pulse toward a target object and measuring a round-trip time of the laser pulse. This measurement principle enables laser scanners to stably offer high-precision scanning abilities even in a large environment. In addition, a noteworthy characteristic of laser scanner is that they obtains a power of the reflected laser (reflectivity) as a side-product of the range data. Laser reflectivity indicates an intensity on the surface of target objects under a single-frequency light source, and a reflectance image, which is a collection of laser reflectance depicted as a grayscale image, contains rich appearance information about the target object. Therefore, laser scanners perform range and intensity measurements for each point simultaneously, and geometric and appearance information principally aligned together is available requiring no calibration setting which is generally essential for a range and image sensor combination.

A laser scanner can be regarded as a single-frequency point light source, and a reflectance image represents diffuse reflection under a point light illumination observed from the sensor position itself. Here, we describe how light reflects and how a reflectance image is captured. Basically the following two factors determine the radiance reflected by a certain point of the scene surface.

- The amount of illumination falling on a surface of the scene
- The amount of the incident illumination reflected by the surface point

To represent the relationship between the incident and reflected light, Bidirectional Reflectance Distribution Function (BRDF) can be used. BRDF of a surface point p, denoted f_p, is defined as the ration of outgoing radiance $L_{p,o}$ and incident irradiance E_p at p as follows:

$$f_p(\omega_i, \omega_o) = \frac{dL_{p,o}(\omega_o)}{dE_p(\omega_i)} \tag{1.11}$$

where $\omega_i \in \Omega$ denotes a solid angle oriented at an angle i. Radiance $dL_{p,i}$ and irradiance dE_p have the following relationship by definition.

$$\begin{aligned} \int_\Omega dE_p(\omega_i) &= \int_\Omega dL_{p,i}(\omega_i) \cos\theta_i d\omega_i & (1.12) \\ \Longleftrightarrow dE_p(\omega_i) dS &= dL_{p,i}(\omega_i)(dS \cos\theta_i) d\omega_i & (1.13) \end{aligned}$$

Based on Eqs. (1.11) and (1.13), the outgoing radiance $L_{p,o}$ is formulated as follows:

$$\begin{aligned} L_{p,o} &= \int_\Omega dL_{p,o}(\omega_i, \omega_o) & (1.14) \\ &= \int_\Omega f_p(\omega_i, \omega_o) dE_p(\omega_i) & (1.15) \\ &= \int_\Omega f_p(\omega_i, \omega_o) L_{p,i}(\omega_i) \cos\theta_i d\omega_i & (1.16) \end{aligned}$$

Assuming the Lambertian reflectance model where the surface luminance is isotropic, the BRDF is simply a constant as $f_p = \rho$. In addition, considering rectilinear propagation of laser and light decay with distance r, a reflectance value $R_{p,o}$ is represented as follows:

$$\begin{aligned} R_{p,o} &= \frac{1}{r^2} \int_\Omega \rho L_{p,i}(\omega_i) \cos\theta_i d\omega_i & (1.17) \\ &= \frac{\rho L \cos\theta_i}{r^2} & (1.18) \end{aligned}$$

where L denotes the power of emitted laser form a laser scanner, and ρ denotes diffusion coefficient of target object's surface. According to Eq. (1.18), a reflectance image is composed of by the following elements:

- Power of emitted laser pulse
- Reflectance property of the surface of the target object
- Relative angle between the incident laser pulse and the surface normal

It is also considered to be affected by sensor specific characteristics. Shinozaki et al. formulated the lase reflectance including the variation in sensitivity of a sensor unit $k(r)$ as follows[19]:

$$R_{p,o} = k(r)\frac{\rho L \cos\theta_i}{r^2} \tag{1.19}$$

1.4 Aim

Time-of-flight laser scanner is one of the most efficient range sensors which can provide accurate 3D range information stably without being affected by disturbance light. The aim of this dissertation is to enhance the functionality of time-of-flight laser scanners for further accurate modeling and understanding of the 3D world.

In order to model an object with high realism, not only geometry but also appearance (texture) information is essential. The 3D geometric model and texture image are generally captured by different sensors, such as a laser scanner and a digital camera, respectively. Therefore, strict calibration between these sensors is indispensable to assign color information to the surface of the 3D geometric model, however, even small alignment errors cause unsatisfactory gaps between the geometric model and the assigned texture.

On the other hand, as discussed above, laser scanners can provide a power of the reflected laser as a side-product of the range data. A reflectance image, which is a collection of laser reflectivity depicted as a grayscale image, indicates intensity on the surface of a target object. Additionally, a unique reflectance value is determined for each pixel in the corresponding range image. Thus, laser scanners are able to measure geometric and appearance information principally aligned together.

Despite of the advantage of the modality, few relevant works have exploited the attribute of laser reflectivity. As rare examples where the reflectance information is used, Kurazume et al.[20] and Inomata et al.[21] developed high-accuracy texture mapping techniques by estimating relative pose and position between reflectance and camera images. Shinozaki et al.[19] proposed a high-quality texturing technique by correcting input camera images focusing on that the diffuse component of the scanned 3D model can be extracted from the reflectance image. These studies provided efficient alignment or color correction approaches for texture mapping using reflectance images, however, they were no more than mediating the 3D model and input texture images through laser reflectivity.

As discussed in the previous section, since a reflectance image is composed of several elements such as incident angle, distance, and diffuse component, much information can be extracted from it which cannot be seen in the corresponding range image. Focusing on the attribute of the laser reflectivity, this dissertation develops up a new research area of 3D modeling with a laser scanner by **complementarily** exploiting range and reflectance information.

1.5 Contributions

Time-of-flight laser scanners provide accurate 3D geometric information stably without being affected by lighting condition, and play important roles in a variety of research fields. In recent years, high-precision three-dimensional laser scanners, such as RIEGL VZ-400 (RIEGL GmbH), Leica Scan Station 2 (Leica Geosystems AG), and TOPCON GLS-1500 (TOPCON), have been widely used for landscape surveying or digital 3D modeling. In addition, a low-cost, high-resolution laser measurement systems using two-dimensional laser scanners (SICK LMS151 (SICK AG) and HOKUYO TOP-URG (HOKUYO)) and a rotary table have been developed for 3D environmental map building for mobile robot navigation[4]. These sensors can acquire high-resolution and precise range data, however, they also have some drawbacks: Due to the measurement principle, laser scanners often suffer from reflectance properties of objects' surfaces and range images acquired with laser scanners include noise and holes. They also cannot capture color textures of the target objects.

The goal of this dissertation is to enhance functionality of time-of-flight laser scanners for better 3D modeling, and enable them to **reconstruct properly colored/textured 3D models with little measurement error and no holes** by utilizing laser attribute. This dissertation consists of four main topics:

In Chapter 2, a novel range image denoising technique exploiting the corresponding reflectance image is proposed. While laser scanners are able to capture precise range images, the range images are generally corrupted by noise because of the electrical and mechanical disturbances or the surface reflectance properties of target objects. To deal with this issues, a novel smoothing filter that refines deteriorated range images taking laser reflectance into consideration is developed.

In Chapter 3, we developed a completion technique which restores ”holes” in a range image referring the corresponding reflectance image. Laser scanners sometimes fail to measure range data because of specular reflection or weak reflectivity of the laser pulse. This results in missing data that appears in the range image in the form of ”holes”. In order to restore the deteriorated range image, we proposed a novel range image completion technique in optimization framework with belief propagation considering the continuity of range and reflectance data.

Inspired by image colorization, a novel 3D geometric colorization method is proposed in Chapter 4. The proposed method colorizes the reflectance image directly using several color cues from an uncalibrated camera image taken in the same scene. This enables us to give color information to 3D geometric models scanned with laser scanners requiring no strict calibration which is essential for the conventional texture mapping.

In Chapter 5, a new texture synthesis technique for hole-free texture mapping is proposed. Texture mapping enables us to assign color images to the target geometric data and to create a realistic virtual 3D model. However, in case that the positions of the camera and the laser scanner are different, some texture-less regions (holes) may exist on the object surface where the appearance information is missing due to the occlusion or out-of-sight of the camera. Considering that completion order significantly affects the quality of the synthesized image, the proposed technique completes the holes by synthesizing texture according to structural information extracted from corresponding reflectivity.

2

Range Image Smoothing using Trilateral Filter

Acquiring 3D geometric information in the real world is indispensable for many applications in robotics, computer graphics, and computer vision. Typical examples include mobile robot localization, augmented reality, human-computer interaction, gaming, and so on. Over the last decades, time-of-flight laser scanners, that are capable of measuring accurate 3D geometric information stably without being affected by lighting condition, have been playing important roles for application such as landscape surveying and cultural heritages of famous buildings/art works. However, the range images captured with laser scanners often suffer from noise due to the reflectance property of objects' surfaces or electrical and mechanical disturbances. For example, one sigma accuracy of RIEGL VZ-400 (RIEGL GmbH) is 3mm per 100 meters, thus a flat surface is measured as a slightly uneven plate. Therefore, denoising techniques for range images taken by laser scanners still remains as a critical problem.

Several approaches can be used to denoise range images captured by range sensors:

1. Averaging a series of range images of the same scene (temporal smoothing),
2. Applying spatial smoothing filters, such as Gaussian filter (spatial smoothing), [22],[23],[24],[25],[26]
3. Combining range data with other information, such as texture or brightness[27],[28],[29],[30]

Temporal smoothing is an intuitive and fundamental technique for denoising a range image and is widely used in high-precision laser scanners. However, this process must be performed for each measured point (pixel) during the measurement, and the processing time is proportional to the number of images for averaging.

On the other hand, spatial smoothing can be applied off-line and is applicable to not only measured points but also structured meshes. In this technique, range values of adjacent points (pixels in a range image or vertexes in structured meshes) are spatially convolved by applying a spatial convolution filter, such as a median filter or a Gaussian filter.

In this chapter, we develop a new denoising technique, which can be categorized into the third category mentioned above by focusing on the laser reflectivity. When we measure range data with laser scanners, the laser reflectivity which indicates the strength of the reflected light can be obtained as a by-product of range data. Note that all of the pixels in the range image have corresponding reflectance values. In other words, the range image and the reflectance image are precisely and fundamentally aligned.

Using the reflectance image, we propose a new smoothing technique using the trilateral filter and the laser reflectivity. In the proposed method, the trilateral filter is applied for not only a range image but also the corresponding reflectance image. By taking account of the properties of range and reflectance images, the proposed method can smooth a range image while preserving geometric features such as jump and roof edges. Adelsberger et al.[31] proposed a similar denoising technique for low-resolution infrared time-of-flight sensors (SwissRanger, MESA-Imaging). However, detailed discussion for the performance has not been presented.

In Section 2.1, an overview of the previous approaches will be presented. In Section 2.2, we will describe the algorithm of the proposed denoising technique for range images using reflectance images, that is, range image smoothing by the trilateral filter. In Section 2.3, simulations and experiments using laser scanners will be reported for the purpose of verifying the performance of the proposed technique.

2.1 Related work

Smoothing techniques for range images are classified into two categories: pixel-based or point-based techniques [26],[27],[32],[33],[29],[30] and mesh-based techniques [23],[24],[25],[28]. Raw range data acquired by range sensors is composed of a group of 3D points called a point-cloud. Pixel or point-based methods denoise the range image or the point-cloud directly without taking the continuity of pixels into account explicitly. On the other hand, mesh-based methods are applied to structured meshes, such as triangular patches, by considering the continuity of the vertexes in the structured meshes.

For the case in which a high-resolution gray-scale image and a low resolution range image are simultaneously captured from a range sensor, Diebel et al.[27] proposed a technique for estimating high-resolution range images by considering the Markov Random Field in high- and low-resolution range images and adjusting smoothing parameters according to the gradient of the high-resolution gray-scale image. Crabb et al.[34]

and Chan et al.[32] also proposed an up-sampling technique using the joint bilateral filter[35]. Bohme et al.[28] proposed a denoising technique for range images using the shape-from-shading technique[36]. They introduced an energy function consisting of the difference of the observed laser reflectance and its estimation based on the Lambertian reflectance model and the continuity of the range and reflectance images. Then, the energy function is minimized by the non-linear conjugate gradient method so that the noise in the range image is suppressed. Using measured range information and the normal directions from photometric stereo, Nehab et al.[29] and Okatani et al.[30] proposed denoising techniques for reconstructing 3D geometric models precisely. Nehab et al.[29] refines the bias in the measured normal direction at first, and then optimizes the 3D model so that the estimated normal direction fits the 3D model. Okatani et al.[30] estimates the shape of an object by integrating the surface normal and 3D data in probabilistic framework.

On the other hand, several techniques based on the bilateral filter [37], which was developed as an edge-preserving filter for gray-scale images, have been proposed[24],[25],[26],[38]. Fleishman et al.[25] proposed a 3D edge preserving filter by applying the bilateral filter for the distance from a point to its adjacent points projected on a tangential plane (tangential component) and the distance from the adjacent points to the tangential plane (normal component). Jones et al.[24] proposed a similar technique using triangular meshes instead of tangential planes. However, these smoothing techniques are applied after converting the point-cloud to the meshes and it is difficult to obtain the normal vectors stably from meshes that contain a great deal of noise. Moreover, in some cases, the construction of structured meshes from a noisy point-cloud is not a simple and trivial problem.

Miropolsky[26] proposed a geometric bilateral filter, which uses the distances from the adjacent points and the difference of normal directions for each point in the point-cloud. However, a stable solution of normal vectors from a noisy point-cloud has not yet been found.

While these methods can be considered as a simple extension of the bilateral filter for a gray-scale image to a range image, the technique proposed herein uses a reflectance image whose pixels correspond one-to-one to those of a range image for denoising the range image. Note that although we assume that the range and reflectance images have the same resolution, the proposed method can be applied to images having different resolutions using the joint bilateral filter[35]. On the other hand, similar denoising techniques for camera images that extend the bilateral filter to other domains have been proposed[39][40][41][42]. Eisemann et al.[39] and Petschnigg et al.[40] proposed flash photography techniques. In their techniques, an image, which was taken under dark illumination and suffers from noise or blur, is reconstructed using itself and another image taken in the same position with flash light. Focusing on the fact that the flash image has better high-frequency information than the no-flash image, they proposed Cross-bilateral filter and Joint bilateral filter respectively. These filters had edge-preserving weighting functions based on the differences of pixel intensities in the flash image in order to obtain the detail shapes. Benett et al.[41] and Weber et al.[42] proposed similar bilateral filtering techniques for each image in a video and animation. They extended the bilateral filter to time domain and denoised each image utilizing the information of adjacent frames.

2.2 Range Image Smoothing using Trilateral Filter

In this section, we propose a new technique for smoothing a range image using the trilateral filter and a reflectance image. As mentioned above, conventional smoothing techniques for range data are mainly applied for a range image directly. On the other hand, we focus on a reflectance image that is acquired as a by-product of the range image for most time-of-flight laser scanners. By taking the properties of both range and reflectance images into account, the proposed technique can suppress noise in the range image while preserving geometric features such as jump and roof edges.

In the following sections, we introduce the conventional bilateral filter, and then the proposed trilateral filter using a reflectance image is described in detail.

2.2.1 Bilateral filter

The bilateral filter[37] is an edge-preserving smoothing filter that extends the Gaussian filter so that not only the spatial relation but also the variation of the pixel intensity is considered. In the Gaussian filter, it is assumed that adjacent pixels may have similar intensities, and a weighted sum of neighbor pixels defined by a Gaussian distribution is calculated for each pixel as follows:

$$g_i = \frac{\sum_{j \in S_i} w_x(x_i, x_j) f_j}{\sum_{j \in S_i} w_x(x_i, x_j)} \tag{2.1}$$

$$w_x(x_i, x_j) = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{|x_i - x_j|^2}{2\sigma_x{}^2}} \tag{2.2}$$

where x_i is the location of pixel i, g_i is a smoothed intensity value of a pixel i, and f_j is the original intensity value of a pixel j, which is a neighbor S_i of a pixel i. Here, $w_x(x_i, x_j)$ is the weight function determined by a Gaussian function with a variance of σ_x^2, and $|x_i - x_j|$ is the two-dimensional spatial distance (L2 distance) between pixels i and j.

Compared with the Gaussian filter, the bilateral filter takes the variation of the pixel intensities into consideration for image smoothing. More precisely, the bilateral filter determines a weight of each neighbor pixel according to not only the two-dimensional spatial relation but also the similarity of the intensity, as follows:

$$g_i = \frac{\sum_{j \in S_i} w_x(x_i, x_j) w_f(f_i, f_j) f_j}{\sum_{j \in S_i} w_x(x_i, x_j) w_f(f_i, f_j)} \tag{2.3}$$

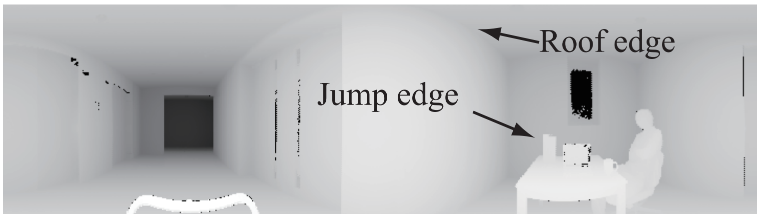
$$w_f(f_i, f_j) = \frac{1}{\sqrt{2\pi\sigma_f}} e^{-\frac{|f_i - f_j|^2}{2\sigma_f{}^2}} \tag{2.4}$$

where $w_f(f_i, f_j)$ is the weight function for the intensity determined by a Gaussian function with a variance of σ_f^2, and $|f_i - f_j|$ is the difference in intensity of pixels i and j. Since the bilateral filter takes the difference in intensity into account, it is possible to preserve abrupt changes in range images, such as jump edges, which are blurred by

Gaussian filter.

2.2.2 Trilateral filter

Bilateral filter is effective to suppress noises in range images while preserving their geometric features. However, as shown in Fig.2.1(a), although abrupt changes of range values, such as jump edges, are easily detected in a range image, moderate changes, such as roof edges, are quite difficult to detect. Miropolsky[26] introduced the directional variation of normal vectors in order to emphasize these moderate changes in the range image. However, if we see the reflectance image shown in Fig.2.1(b), these moderate changes are clearly detected in the reflectance image.



(a) Range image

(b) Intensity image

Figure 2.1: Jump and roof edges in range and reflectance images

Based on the above consideration, we propose a new filter that uses reflectance and range images simultaneously for smoothing a range image, as follows:

$$g_i = \frac{\sum_j w_x(x_i, x_j) w_f(f_i, f_j) w_d(d_i, d_j) f_i}{\sum_j w_x(x_i, x_j) w_f(f_i, f_j) w_d(d_i, d_j)} \tag{2.5}$$

$$w_x(x_i,x_j) = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{|x_i-x_j|^2}{2\sigma_x{}^2}} \tag{2.6}$$

$$w_f(f_i,f_j) = \frac{1}{\sqrt{2\pi\sigma_f}} e^{-\frac{|f_i-f_j|^2}{2\sigma_f{}^2}} \tag{2.7}$$

$$w_d(d_i,d_j) = \frac{1}{\sqrt{2\pi\sigma_d}} e^{-\frac{|d_i-d_j|^2}{2\sigma_d{}^2}} \tag{2.8}$$

where f_i and d_i are the range and reflectance values in pixel i, and $w_x(x_i,x_j)$, $w_f(f_i,f_j)$, and $w_d(d_i,d_j)$ are Gaussian functions in the two-dimensional spatial, range, and reflectance domains with variances of σ_x^2, σ_f^2, and σ_d^2, respectively.

The filter given by Eq. (2.5) takes into account three kinds of information in range and reflectance images for smoothing a range image. In other words, it is an extension of the conventional bilateral filter that takes the variation of the laser reflectivity into consideration for range image smoothing. Thanks to a variety of properties in range and reflectance information, the proposed trilateral filter enables not only jump edges but also roof edges to be preserved in a range image, and the trilateral filter is expected to have higher performance for edge preservation than the simple extension of the bilateral filter to a range image.

Consequently, the proposed smoothing technique for a range image is summarized as follows.

1. Acquire range and reflectance information by a time-of-flight range sensor.

2. Create range and reflectance images in which the values of each pixel in range and reflectance images are proportional to the range and reflectance values.

3. Apply the trilateral filter given by Eq. (2.5) using range and reflectance images and obtain a smoothed range image.

4. Construct a 3D model consisting of meshes from the smoothed range image.

2.3 Results

This section introduces the results of the range image smoothing experiments using simulated and actual range images. We conducted experiments with various parameters selected manually, and determined parameters used for the following experiments.

2.3.1 Simulation using a synthesized image

First, we performed simulation experiments using the synthesized image shown in Fig.2.2, which is a scene of a square box having sides of 1 meter in a room. Since a reflectance image indicates an appearance of a target object under a single-frequency light source, the gray-scale image (Fig.2.2(a)) is synthesized as an alternative to it by placing a point light source at the same position as a virtual viewpoint. We added random noise of 1 % of a range value of each pixel in the range image.

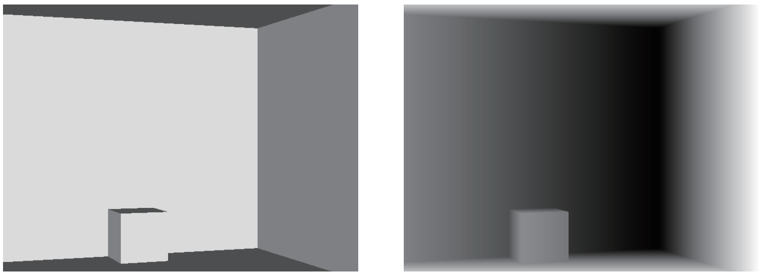
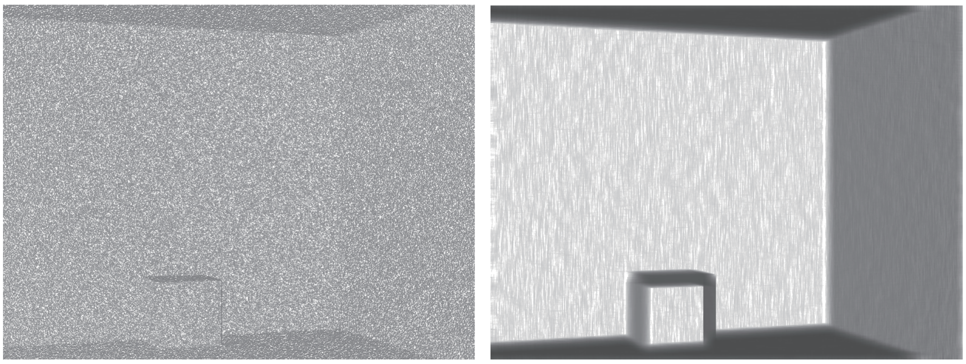

(a) Grayscale image (b) Range image

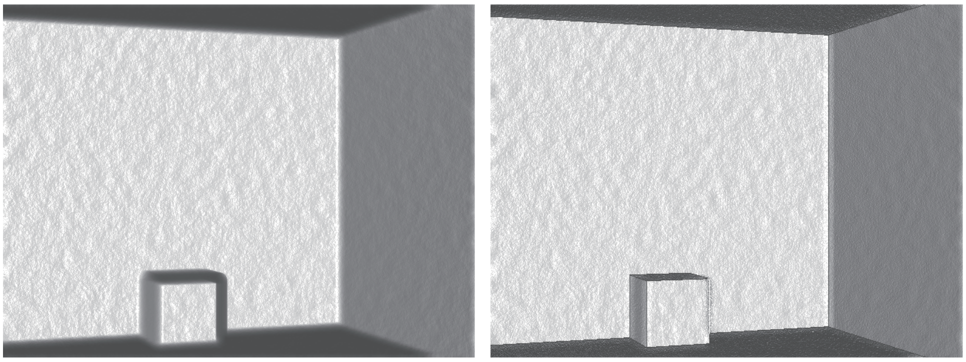
Figure 2.2: Synthesized images used in the simulation experiment

Figure 2.3 shows the denoising results obtained using the Gaussian filter, the bilateral filter, and the trilateral filter, respectively. Table 2.1 shows the RMS errors of the range images after applying these filters.

In this experiment, the kernel size of each filter is 9×9 pixels, and the ranges of the range data and the reflectance data are 13,293 to 17,128[mm] and 0 to 255, respectively. The variances are set as $\sigma_x = 4.0$, $\sigma_f = 0.4$, and $\sigma_d = 6$ for the normalized range image. Computational times to apply these filters to the range image whose size is 640×480 pixels with Intel Xeon (2.50GHz) were 202ms for Gaussian filter, 580ms



(a) 1% noise in depth values (b) Gaussian filter



(c) Bilateral filter (d) Trilateral filter

Figure 2.3: Denoised images by Gaussian filter, bilateral filter and trilateral filter

for bilateral filter, and 824ms for trilateral filter.

As shown in Table 2.1, the RMS error of the proposed trilateral filter is the smallest, and it is verified to have high performance for range image smoothing and edge preservation.

Additionally, assuming the reflectance image also contained noise, we carried out another experiment using a synthesized reflectance image created by adding 1 % random noise to Fig.2.2(a). In this experiment, we adopted the following three methods and compared their performances.

1. Merely apply the proposed trilateral filter to the range image with the noisy reflectance image.

2. Apply Gaussina filter to the noisy reflectance image, and then denoise the range

	RMS [mm]
Original image	45.8
Gaussian filter	17.8
Bilateral filter	14.1
Trilateral filter (proposed)	11.7

Table 2.1: RMS error for range image smoothing

image with the smoothed reflectance image.

3. Apply bilateral filter to the noisy reflectance image, and then denoise the range image with the smoothed reflectance image.

Table 2.2 shows the RMS error after applying each method. This demonstrates that the proposed trilateral filter works well even with a noisy reflectance image, and that applying smoothing filter to a noisy reflectance image beforehand enhances the performance of the trilateral filter to some extent. Here, the kernel size and variances of smoothing filters for the reflectance image are set as 11×11, $\sigma_x = 4$, and $\sigma_f = 7$.

Table 2.2: RMS error (1% noise for range and reflectance images)

	RMS [mm]
Original image	45.8
Trilateral filter without smoothing	20.5
Trilateral filter after applying Gaussian filter	19.4
Trilateral filter after applying bilateral filter	19.1

2.3.2 Experiments with a laser scanner

Next, we performed experiments using a 3D laser measurement robot CPS-V shown in Fig.2.4[4]. Figure 2.5 shows three experimental conditions: a simple environment consisting mainly of roof edges (scene 1), a more complex environment in which a human and other objects exist and a number of jump edges are observed (scene 2), and a stone monument with a Kanji inscription which contains roof and jump edges (scene

3). Figure 2.6 shows the range and reflectance images of these scenes captured by the measurement robot. The robot enables the surrounding range data to be captured by rotating the laser scanner (SICK, LMS151 or LMS200) by means of a rotary table. The image sizes are 200 × 721 pixels for scene 1 and 2, and 760 × 1133 pixels for scene 3. Figure 2.6 displays partial views of these range and reflectance images.

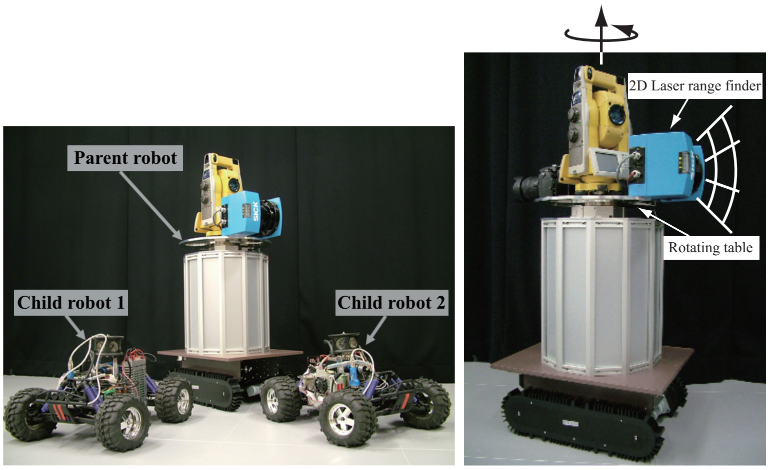


Figure 2.4: Acquisition system of a panoramic range image[4]

Figure 2.7 shows the results for scene 1 (Fig.2.5(a)). In the experiment, we set the kernel size of the filters to be 9 × 9 pixels, and the ranges of the range data and the reflectance data are 275 to 8,191 [mm] and 0 to 255, respectively. The variances are $\sigma_x = 0.8$, $\sigma_f = 0.1$, and $\sigma_d = 7$ for the normalized range image.

Figure 2.7(a) is a 3D model constructed from the original range image before applying smoothing filters. Several unexpected bumps appear on the surfaces of the walls and objects due to the noise in the range image. Figures 2.7(b), 2.7(c), and 2.7(d) show the images smoothed by the Gaussian filter, the bilateral filter, and the trilateral filter, respectively. These figures show that the surfaces of the walls are smoothed by the Gaussian filter and the bilateral filter. However, the edges of the box and the window frame are blurred. On the other hand, the trilateral filter can smooth the surfaces of the walls while preserving the edges of the box and the window frame.

Figure 2.8 displays comparison of a cross-sectional shape of the box in a 3D model

(a) Scene 1 (b) Scene 2

(c) Scene 3

Figure 2.5: Experimental setup

illustrated in Fig.2.7(a) after applying each filter. While Gaussian filter and bilateral filter over-smooth and blur the prominent structure, the proposed trilateral filter maintains the geometric structure suppressing the noise. Here, processing time of each filter with Intel Xeon (2.50GHz) was 104ms for Gaussian filter, 276ms for bilateral filter, and 414ms for trilateral filter.

Next, the results for scene 2 (Fig.2.5(b)) are shown in Fig.2.9. Figure2.10 displays an enlarged partial view of each result. Similar to the experimental results for the scene 1, the Gaussian filter and the bilateral filter smooth the surfaces of the walls. In particular, the bilateral filter preserves the jump edges, such as the shape of the monitor, which is indicated by the arrow. However, the roof edges of walls, the face of the person, and the ruck of the clothes are blurred. On the other hand, the trilateral filter smoothes the range image successfully while preserving the jump and roof edges appropriately, as shown in Fig.2.9(d). In this experiment, we used the same parameters as in scene 1.

Finally, the results for the scene 3 (Fig.2.5(c)) are shown in Fig.2.11. In this experiment, we set the kernel size of the filters to be 9×9 pixels, and the ranges of the range data and the reflectance data are 443 to 49,726 [mm] and 0 to 255, respectively. The variances are $\sigma_x = 3$, $\sigma_f = 0.3$, and $\sigma_d = 13$ for the normalized range image. As shown in Fig.2.11(b), the bilateral filter blurs the edges of the characters while smoothing the surfaces of the stone monument. Figure 2.11(c) shows the result after applying Bohme's technique[28] that estimates a Lambertian reflection coefficient and a normal direction on each part of the surface simultaneously by minimizing an energy function. Since this technique utilizes the difference between reflectivity and a normal direction in each pixel and their averages of its neighbor pixels as a smoothing term, the obtained range image also fails to conserve the detailed geometric features as is the case with the mean filter and the Gaussian filter. On the other hand, the trilateral filter preserves the geometric features successfully suppressing noises in the range image, as shown in Fig.2.11(d).

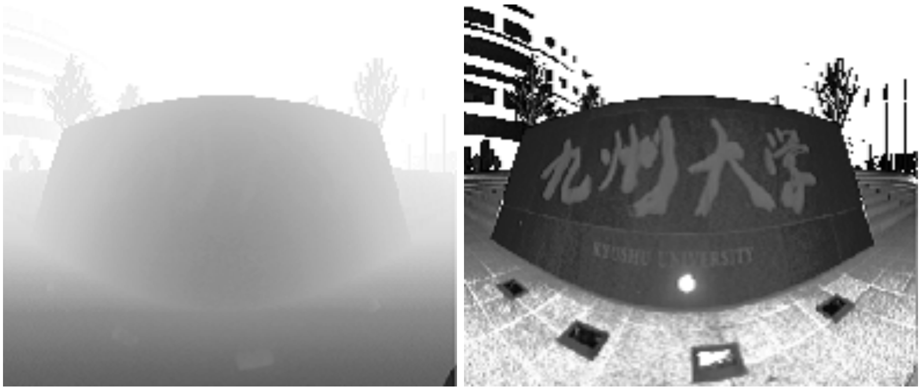
Computational times to apply these filters to the range image whose size is 640×480 pixels with Intel Xeon (2.53GHz) were 347ms for Gaussian filter, 1361ms for bilateral filter, 2574531ms for Bohme's technique, and 1994ms for trilateral filter.

Range image Reflectance image

(a) Scene 1

Range image Reflectance image

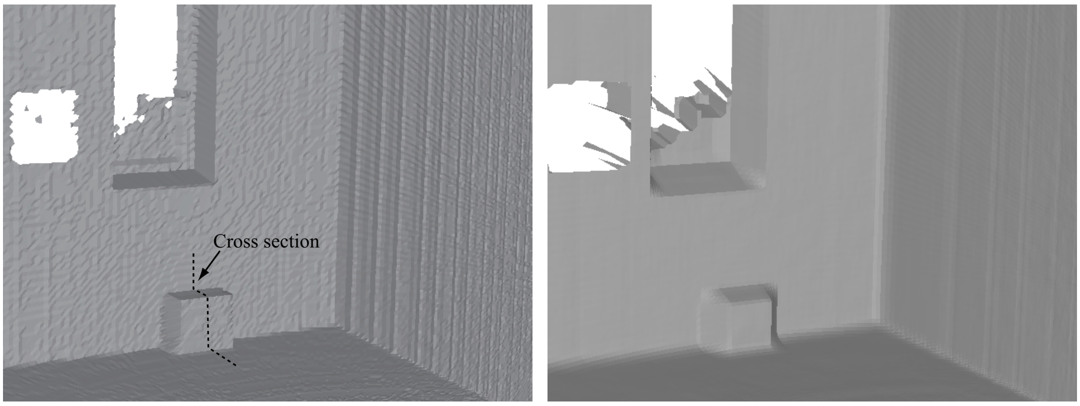
(b) Scene 2



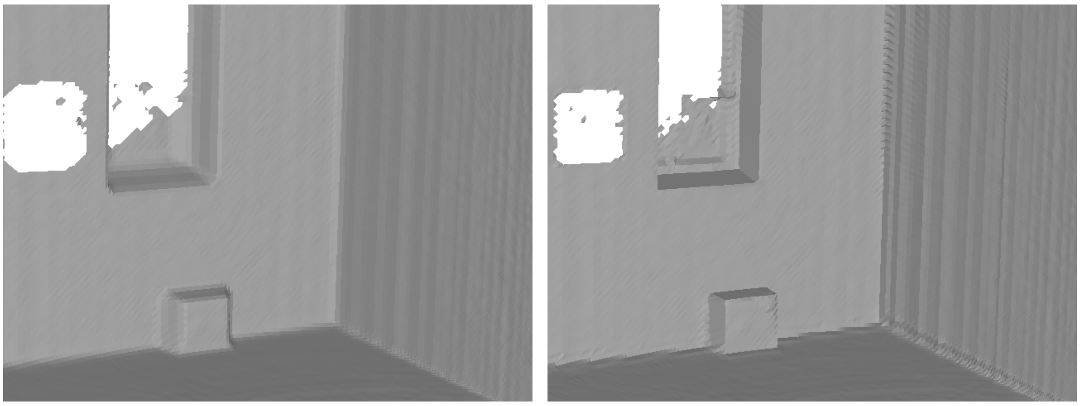
Range image Reflectance image

(c) Scene 3

Figure 2.6: Range and reflectance images

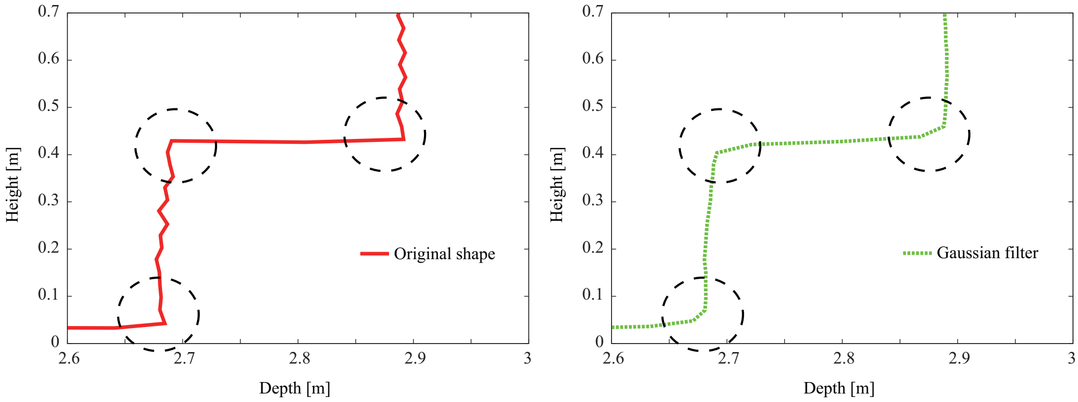


(a) Original range image (b) Gaussian filter

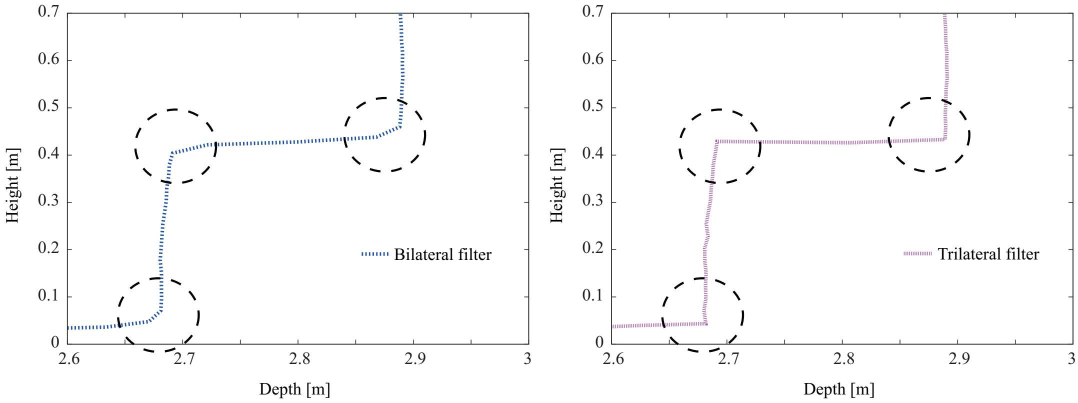


(c) Bilateral filter (d) Trilateral filter

Figure 2.7: Experimental results for a simple environment



(a) Original shape (b) Smoothed shape by Gaussian filter

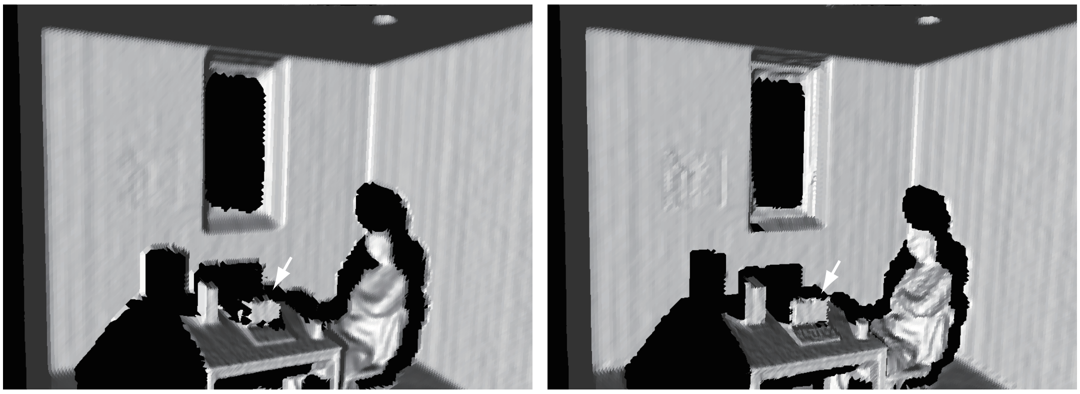


(c) Smoothed shape by bilateral filter (d) Smoothed shape by trilateral filter

Figure 2.8: Comparison of cross-section shape of the box

(a) Original range image

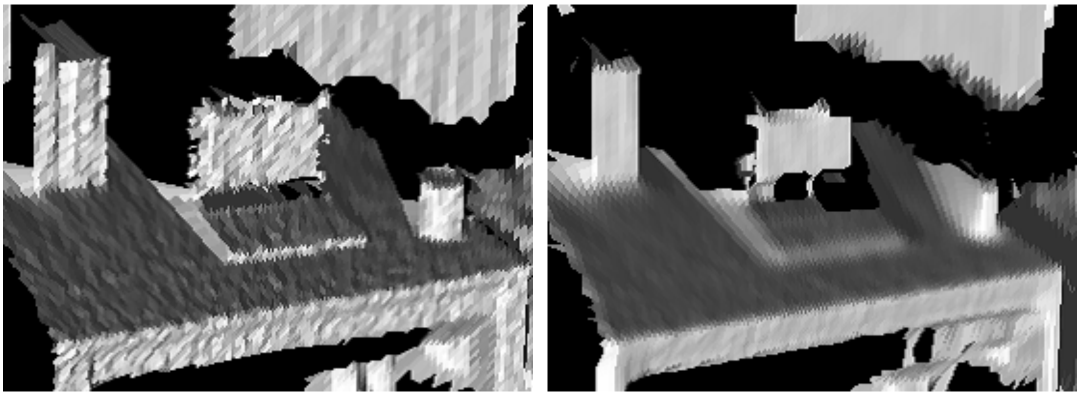
(b) Gaussian filter



(c) Bilateral filter

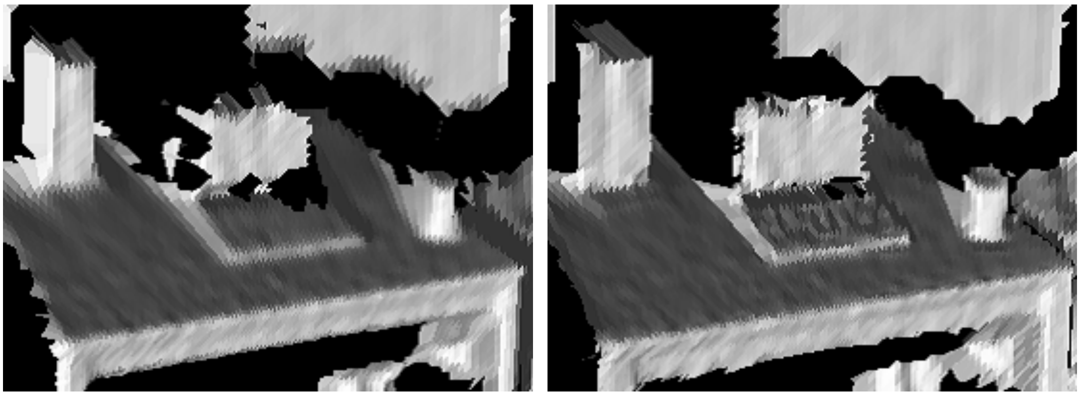
(d) Trilateral filter

Figure 2.9: Experimental results for a complex environment



(a) Original range image

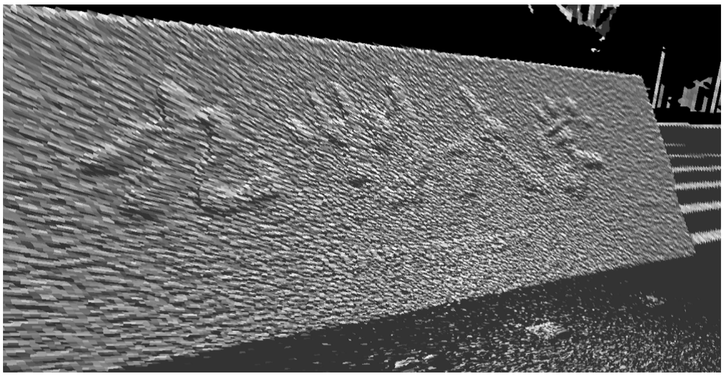
(b) Gaussian filter



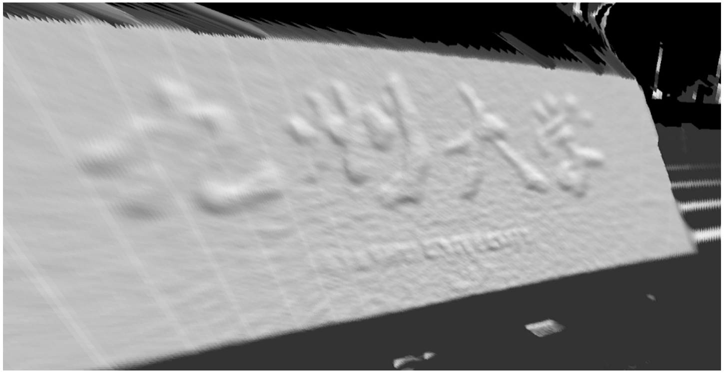
(c) Bilateral filter

(d) Trilateral filter

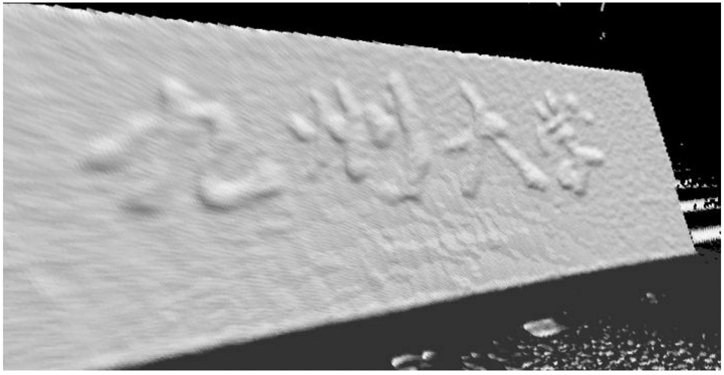
Figure 2.10: Objects on the table



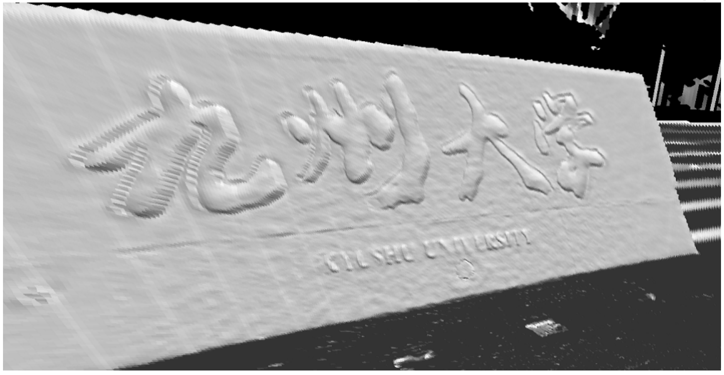
(a) Original range image



(b) Bilateral filter with range image



(c) Bohme's method[28] with range and reflectance images



(d) Trilateral filter with range and reflectance images

Figure 2.11: Experimental results for a stone monument

2.4 Conclusions

In this chapter, we proposed the trilateral filter for range image smoothing focusing on the laser reflectivity. By taking into account the properties of range and reflectance images, the proposed trilateral filter can suppress noises in range images while preserving geometric features such as jump and roof edges. We conducted experiments using synthesized images and actual range images and verified that the proposed denoising technique successfully suppress noises in range images.

Since the reflectance image is obtained as a by-product of range data, the proposed method has several advantages. For example, no additional measurements or instruments are required, and, unlike conventional camera images, the reflectance image is not affected by lighting conditions. In addition, while a range image has the advantage of detecting jump edges, a reflectance image is suitable for detecting roof edges. Therefore, it is expected that the proposed technique has higher performance in edge preservation than the techniques that use range or reflectance information only.

In the future, we will discuss the optimum parameters for the proposed technique and perform quantitative evaluation for a variety of scenes.

3

Range Image Completion with Belief Propagation

Time-of-flight laser scanners are capable of measuring highly accurate 3D range information without being affected by disturbance light, and increasingly used for a variety of applications in the fields of robotics, human-computer interaction, autonomous vehicle, architecture, and so on. Laser scanners infer distances to scene surfaces from themselves based on round-trip times of laser pulses emitted to the scene and reflected off the surface. However, depending on the reflectance properties of objects' surfaces, laser scanners fail to measure range data due to specular reflection or weak reflectivity of the laser pulse. For example, metal surface with strong specular reflection or black color cannot be measured by standard laser scanners, and the missing data appears in the range images in the form of holes. Note that, in general, not only the range data but also the reflectance data in those regions is lost.

In this chapter, we propose a new completion technique which restores holes included in a range image utilizing a corresponding reflectance image and belief propagation. Reflectance data tends to be more sensitive to geometric changes than range data, and geometric features such as jump and roof edges can be observed more clearly in a reflectance image. Focusing on the fact, we first restore a deteriorated reflectance image with belief propagation in order to acquire a rough estimate of the appearance. Then, we apply an extended belief propagation to a corresponding deteriorated range image which weights belief messages among adjacent pixels according to continuities of reflectance data in the restored reflectance image. This 2-step completion technique enables us to restore deteriorated range images more appropriately than directly applying belief propagation to them.

The rest of this chapter is organized as follows. We first discuss the previous completion techniques in Section 3.1. In Section 3.2 we describe the detail algorithm of proposed completion technique extending loopy belief propagation utilizing the laser reflectivity. Section 3.3 demonstrates the validity of the proposed technique carrying out simulations and experiments. Finally, we conclude this chapter and mention the future work in Section 3.4.

3.1 Related work

For the case in which there are several holes in range data due to the occlusion or specular/weak reflection, Kawai et al.[43] proposed a completion technique of the 3D surfaces. They define an energy function based on the similarity of shapes, and select the best match which minimizes the energy function to fill in the holes of 3D geometry. Becker et al.[44] proposed a completion method using an additional color image of the same scene from a different viewpoint. The method first creates a lot of small 3D structure patches with the texture from an input 3D geometric model and a texture image aligned to it, and then restores holes in the 3D model by copying the best patch in the set of patches to them. Assuming that regions which have similar textures tend to have similar 3D structures, they select the best patch from the patch collection by comparing texture patterns of patches in the collection and the adjacent area of the hole. Xu et al.[45] also proposed a technique which estimates missing geometry by learning association of surface normals to image patches in calibrated images.

On the other hand, several image completion techniques based on belief propagation have been proposed[46],[47]. Pedro et al.[46] proposed an image completion method which takes the continuity of pixels into account by applying belief propagation. Komodakis et al.[47] addressed exemplar-based image completion as a discrete global optimization problem with belief propagation. They introduced the Priority-BP that extends standard belief propagation for priority-based message scheduling and dynamic label pruning, and that allowed them to avoid visual artifacts and produce excellent completion results while reducing the computational cost of BP.

3.2 Range image completion utilizing 2-step belief propagation

In the previous chapter, we proposed a range image smoothing technique using the trilateral filter and a reflectance image. Although the smoothing technique is able to suppress noises in a range image, a deteriorated range image that has missing parts due to specular reflection or weak reflectivity of the laser pulse is difficult to recover by merely applying the proposed smoothing filter. For restoring the ”holes” in a range image, this chapter proposes an image completion technique using belief propagation and a reflectance image.

3.2.1 Loopy belief propagation

Let us consider a graph P consisting of multiple nodes connected by multiple arcs. We assign label f_p to node p so that the following energy function is minimized.

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{(p,q) \in N} W(f_p, f_q) \tag{3.1}$$

where $D_p(f_p)$ is a cost term for assigning label f_p to node p, and $W(f_p, f_q)$ is a penalty term if labels f_p and f_q are assigned to nodes p and q, respectively. Here, N indicates the neighbor nodes of node p.

In the framework of belief propagation, the following messages are repeatedly exchanged between the adjacent nodes in order to determine the optimum label f_p that minimizes the energy function:

$$m^t_{p \to q}(f_q) = \min_{f_p} \left(D_p(f_p) + W(f_p, f_q) + \sum_{s \in N(p) \backslash q} m^{t-1}_{s \to p}(f_p) \right) \tag{3.2}$$

After T iterations, optimum label f^*_q is determined so as to minimize the following cost function:

$$b_q(f_q) = D_q(f_q) + \sum_{p \in N(q)} m^T_{p \to q}(f_q) \tag{3.3}$$

3.2.2 2-step Belief propagation

We apply belief propagation to a deteriorated range image and repair it using a corresponding reflectance image. When we measure range data using a laser scanner, it often occurs that part of the range image is lost due to saturation of the reflectivity by specular reflection or a weak laser pulse reflected on a black surface. In most cases, not only the range information but also the reflectance information in this region is lost.

The proposed completion technique for a range image consists of two steps. First, we repair a reflectance image by belief propagation in Section 3.2.1, because the reflectance image clearly contains roof and jump edges and the restoration of the reflectance image is easier than the restoration of the range image. Then, we apply belief propagation to the range image using the repaired reflectance image. In Section 3.3, this two-step algorithm is demonstrated to be able to inpaint the range image more precisely than directly applying belief propagation to the range image.

Since belief propagation requires a huge memory and large calculation cost, range and reflectance images are first converted to 256-level gray-scale images. Therefore, the number of labels to be assigned is 256, as expressed by integers from 0 to 255.

We define the cost term $D_p(f_p)$ for assigning label f_p to pixel p as

$$D_p(f_p) = 0 \tag{3.4}$$

for lost regions and

$$D_p(f_p) = \mid f_p - L_p \mid \tag{3.5}$$

for other regions, where L_p is the original label of pixel p. In addition, we consider the four-neighbor q of pixel p and define the cost function for assigning labels f_p and f_q as

$$W(f_p, f_q) = g(r_p, r_q)(f_p - f_q)^2 \tag{3.6}$$

where r_p and r_q are values of pixels p and q in the reflectance image, and $g(r_p, r_q)$ is a gain term that indicates the effect of the reflectance image.

$$g(r_p, r_q) = \alpha e^{-\beta (r_p - r_q)^2} \tag{3.7}$$

Equation (3.6) indicates that the neighboring pixel which has a similar reflectance value is preferentially selected to repair a lost pixel in the deteriorated range image. In contrast, a pixel having a reflectance value that is changed discontinuously affects the repair of the range image only slightly.

3.3 Results

This section shows the results of experiments for range image completion by the proposed belief propagation using simulated and actual range images. We conducted experiments with various parameters selected manually, and determined parameters used for the following experiments.

3.3.1 Range image completion experiments

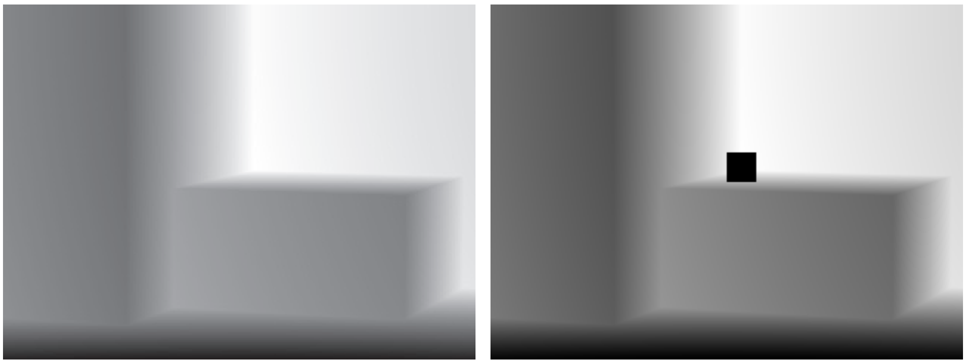
Simulation using a synthesized image

We performed the simulation for evaluating the proposed range image completion technique with belief propagation described in Section 3.2. In the experiment, we prepared deteriorated reflectance and range images which have small missing regions. The size of the image is 320×240 pixels, and the size of the missing part is 20×20 pixels. In this experiment, we use $\alpha = 0.75$ and $\beta = 1.0$.

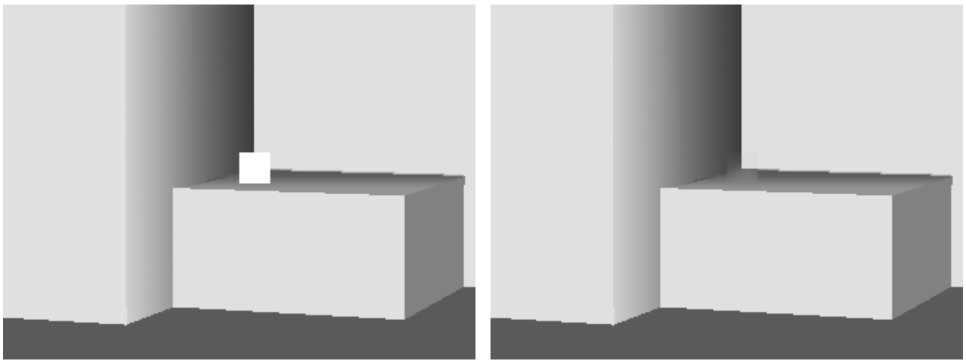
Figures 3.1(a) and 3.1(b) show the original and deteriorated range images, and Figs.3.1(c) and 3.1(d) show the deteriorated and repaired reflectance images. The inpainted range images after applying belief propagation 30 times are shown in Figs.3.1(e) and 3.1(f). These images are repaired with and without the reflectance image respectively, and their 3D mesh models are displayed in Fig.3.2. The RMS errors for these repaired images are compared in Table 3.1. From these results, the range image completion is successfully carried out using the two-step algorithm with belief propagation and the reflectance image.

Experiments with a laser scanner

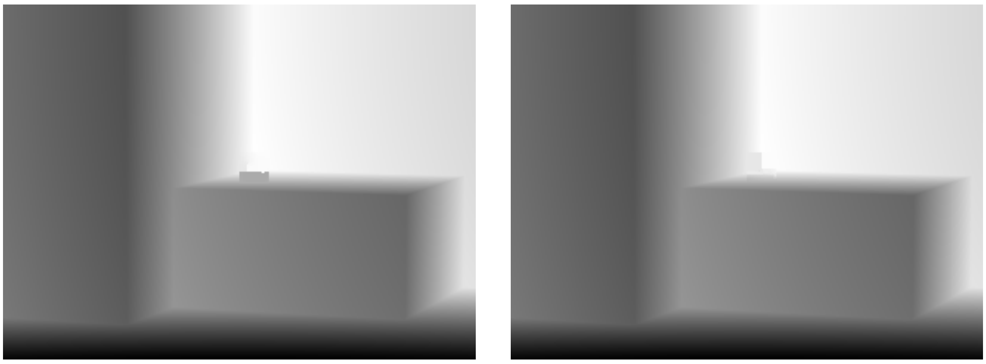
Next, we performed the experiments using actual range and reflectance images taken by the laser scanner on the CPS-V robots (Fig.2.4). In this experiment, we use $\alpha = 0.75$ and $\beta = 0.5$. To emphasise the difference of the restoration results, we prepared deteriorated range and reflectance images by manually cutting a part of scan



(a) Original range image (b)Deteriorated range image



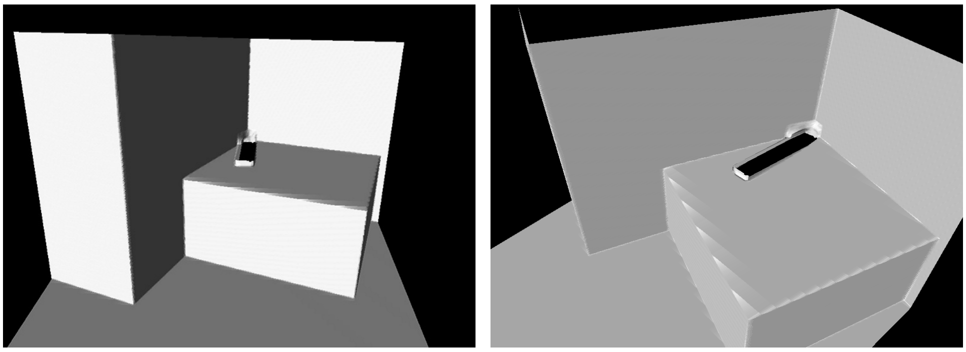
(c) Deteriorated reflectance image (d) Inpainted reflectance image



(e) Completion result with BP (f) Completion result with proposed method

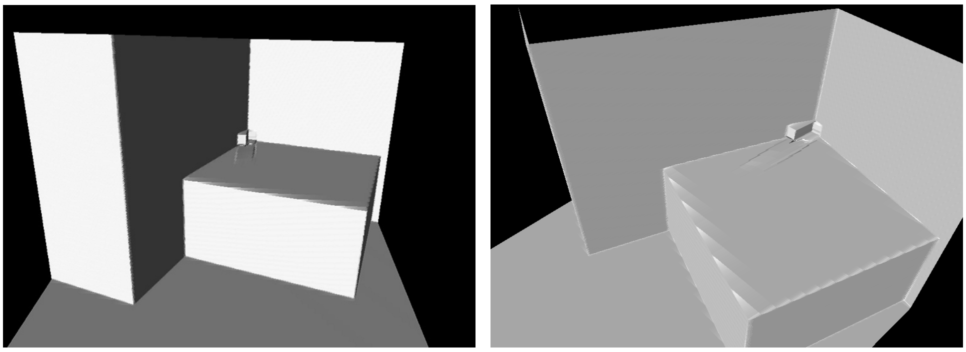
Figure 3.1: Range image completion with Belief Propagation

data. Figure 3.3 shows the 3D models restored by two techniques, that is, the simple belief propagation for range images and the proposed two-step algorithm using laser reflectivity. Missing parts are recovered appropriately by the proposed two-step algorithm while the simple belief propagation produces discontinuities in the restored 3D model. Each RMS error is shown in Table 3.2. As shown in Table 3.2, the RMS error



Front view Side view

(a) Standard BP



Front view Side view

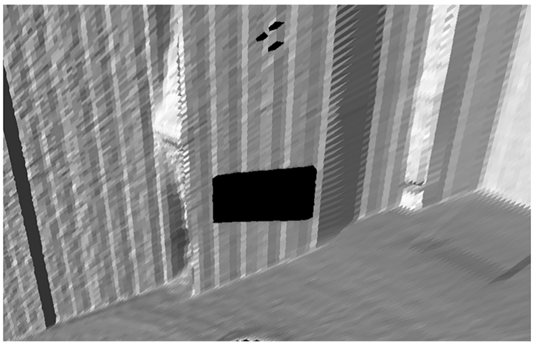
(b) Proposed method

Figure 3.2: Completed 3D mesh model

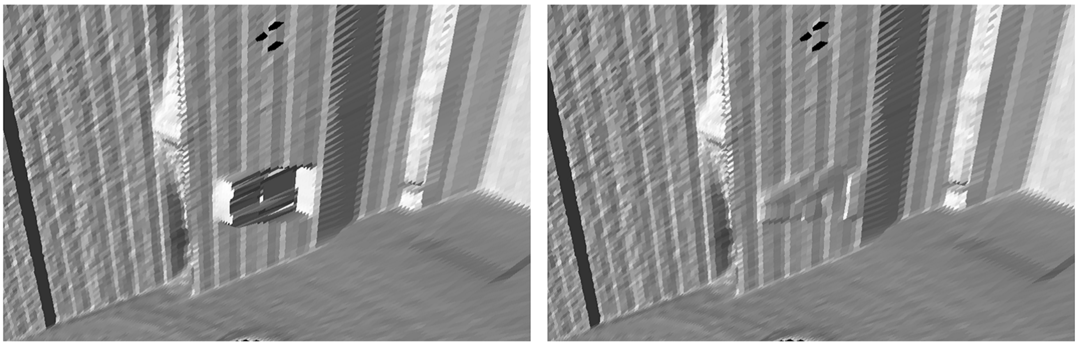
Table 3.1: RMS error for range image completion

Number of iteration	RMS[mm]	
	Without reflectivity	With reflectivity
12	36.24	29.01
20	30.15	14.16
30	28.99	10.68
40	29.47	10.47
50	29.44	10.64

of the proposed technique is the smallest, and the proposed technique is verified to fill-in a hole in the range image successfully.



(a) Original 3D mesh model with a missing region



(b) Standard BP (c) Proposed two-step algorithm

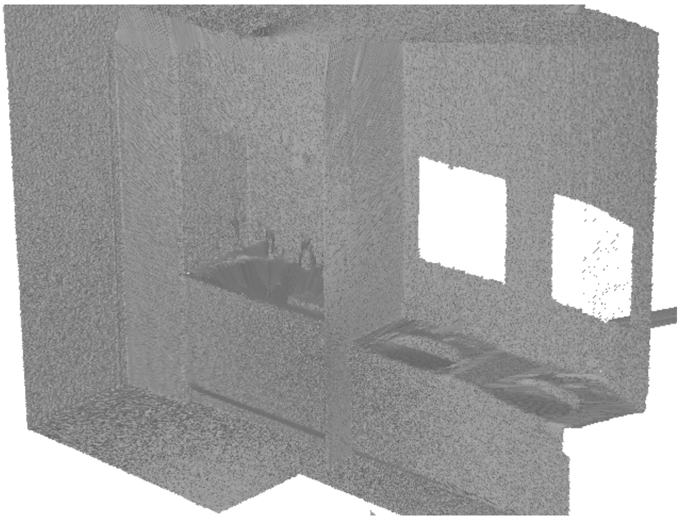
Figure 3.3: Experimental results for the performance evaluation

Table 3.2: RMS error in experimental results for the performance evaluation

	RMS [mm]
Without reflectivity	7.68
With reflectivity (2step)	1.44

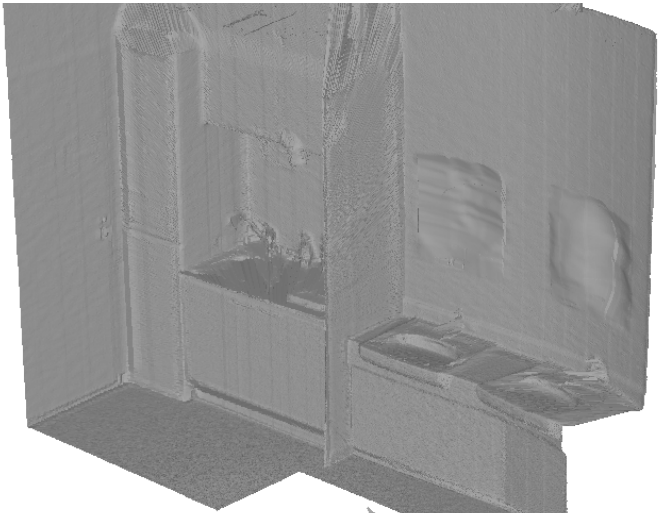
3.3.2 Range image smoothing and completion utilizing laser reflectivity

Finally, we carried out an experiment to verify our smoothing and completion techniques using actual range and reflectance images taken by the laser scanner(LMS511) on the CPS-V robots (Fig.2.4). In this experiment, we set the kernel size of the filters to be 9×9 pixels and the variances are $\sigma_x = 3$, $\sigma_f = 0.3$, and $\sigma_d = 9$ for the trilateral filter. We also use $\alpha = 0.75$, $\beta = 2.0$ and iteration $T = 100$ for our two-step algorithm. Figure 3.4(a) is a 3D model constructed from the original range image. Several unexpected bumps appear on the surfaces of the walls and objects due to the noise in the range image. In addition, several holes can be seen due to specular or weak reflection. Figure 3.4(b) and (c) show the 3D model restored by the two-step completion algorithm, and the smoothed 3D model of it by the trilateral filter, respectively. The two-step algorithm repairs the missing region appropriately by taking the continuity of laser reflectivity into account. Moreover, the trilateral filter smooths the range image successfully while preserving the geometric features such as jump and roof edges.



(a) Original 3D mesh model

(b) The 3D mesh model after applying our two-step algorithm



(c) The 3D mesh model after applying our two-step algorithm and proposed trilateral filter

Figure 3.4: Experimental results for the performance evaluation of proposed smoothing and completion techniques

3.4 Conclusions

In this chapter, we proposed a novel range image completion technique using belief propagation and laser reflectivity. By considering not only the adjacent range values but also the continuity of the reflectance values, the proposed 2-step completion algorithm enables us to recover deteriorated range images more appropriately than directly applying belief propagation to them. We carried out experiments using synthesized data and actual data obtained with a laser scanner respectively, and the results demonstrated the validity of the proposed completion technique.

Currently, the parameters are adjusted manually by conducting experiments with various parameters. Therefore, in the future, we will develop a method that determines optimum parameters automatically according to input data. Besides, we will also improve the proposed technique so that it can restore more complicated geometric structures.

4

Manual/Automatic Colorization for 3D Geometric Models

Three-dimensional geometric modeling of a real object using a laser scanner has been used in many applications such as Virtual Reality (VR), digital archives of cultural heritages[4], and remote control of a rescue robot in a hazardous environment[18]. Moreover, to display a 3D model with high realism, it is effective to add color and texture information to the surface and provide not only a geometry but also appearance information. For instance, it is quite helpful to control a rescue robot in a dark and hazardous environment appropriately if a colored 3D model around a rescue robot is displayed for an operator. Texture mapping[48], which maps a color image obtained by a camera to a 3D geometric model, has been widely used for creating photo-realistic models. However, since color images and the 3D geometric model are generally captured by different sensors, such as a digital camera and a laser scanner, respectively, strict calibration between these sensors is essential to map color information precisely on the 3D geometric model. Moreover, multiple color images are required to add surface appearance to an entire or large-scale 3D model, and small registration errors or changes in lighting conditions cause an unsatisfactory gap and discontinuity due to human perception.

In this chapter, we propose a new technique using laser reflectivity to add color to a 3D geometric model. A laser scanner obtains the range image of a target based on the properties of the laser projected onto the surface. At the same time, the laser reflectivity, which represents the power of the reflected laser, is obtained as a by-product of the distance value. Each pixel in the range image has a corresponding reflectance value. In other words, the range image and the reflectance image are precisely aligned.

By taking advantage of the characteristics of the reflectance image, our proposed technique achieves highly realistic 3D modeling with only a single color image. The key scheme is to colorize the reflectance image based on its similarity with a color image, then transfer the color information to the 3D geometric model. This chapter presents some experimental results using a laser scanner, and quantitative evaluations of the proposed technique by comparing it with texture mapping to demonstrate the validity of the proposed technique.

Whereas previous methods have been developed to assign color information to 3D models by projecting corresponding camera images, we proposes a novel colorization technique for 3D geometric models. The major contribution of our approach is to extend a conventional monochrome image colorization technique so that it can be applied to range and reflectance images. To our best knowledge, no studies have addressed colorization for 3D geometric models scanned with laser scanners by utilizing reflectance attribute except for the proposed method.

The rest of this chapter is organized as follows. In Section 4.1, an overview of the previous approaches will be presented. In Section 4.2, we will propose a new colorization technique for a 3D geometric model using a reflectance image. In Section 4.3, we carry out some experiments using a laser scanner, and verify the performance of the proposed technique.

4.1 Related work

Texture mapping[48], which is a fundamental technique for creating a photo-realistic 3D model, has been widely used in the field of computer vision. In some applications, it maps a color image on a range image and creates a colored 3D model. Usually range and color images are captured from different viewpoints by two independent sensors, such as a laser scanner and a digital camera. Therefore, it is necessary to determine the correspondence between range and color images in order to map color information on the 3D geometric model precisely.

For aligning a 3D model and a color image, Yoshida et al.[49] proposed an alignment technique by assigning several matching points between range and color images manually. Neugebauer et al.[50] proposed a similar technique that calculates suitable camera parameters according to the interactive selection of corresponding points between 3D range data and a 2D color image.

In contrast, techniques that align range and color images automatically have been developed by several researchers. Viola et al.[51] proposed a technique that utilizes statistical characteristics of both images. Stamos et al.[52] also proposed a method that extracts several planes from range data and edges in color images, and then calculates the intersection lines of the planes and the edges. Some approaches that compare a 2D image contour with a silhouette image of the 3D geometric model have been proposed. Iwakiri et al.[53] proposed a real-time texturing method that aligns the color image and the silhouette image of a 3D geometric model from a virtual camera. Lensch et al.[54] also proposed a silhouette-based algorithm that determines the camera transformation based on an XOR-operation between the silhouette image of a 3D model and the color image.

Registration techniques that exploit reflectance images have also been studied. The reflectance image, obtained by most laser scanners as a by-product of the range image, has a similar appearance to a color image. Therefore, some approaches make use of reflectance images for registration between a 2D image and a 3D geometric model for texture mapping. Boughorbal et al.[55] utilized the similarity between the reflectance image and the intensity image based on the χ^2-metric. Umeda et al.[56] proposed a technique to determine relative relations between a range sensor and a color sensor based on the gradient constraint between reflectance and color images. On the other hand, local features in both images are effective to estimate the correspondence between reflectance and color images. Kurazume et al.[20] proposed a calibration method for texture mapping that minimizes the error between edges extracted from reflectance and color images by using the robust M-estimator. Boehm et al.[57] utilized the scale-invariant feature transform (SIFT) [8] to estimate extrinsic parameters by matching them in reflectance and color images. Inomata et al.[21] proposed a SIFT-based technique that calculates not only the extrinsic parameters but also the intrinsic parameter and distortion of the camera lens simultaneously. This method uses Soft-matching that retains correct matches while removing false matches between reflectance and color images according to the similarity of the appearances based on Bhattacharyya distance.

All the methods mentioned above assume that color images which correspond to range images are captured. However, in some cases, it is difficult or almost impossible to provide color images perfectly, for example, in dark and wide areas. The proposed technique in this research gives a solution for providing a colored 3D model even in these conditions.

4.2 Colorization for 3D Geometric Models with Reflectance Images

In this section, we show our algorithm for adding color information to 3D geometric models by applying the image colorization technique to reflectance images [58]. The basic idea behind the proposed technique is as follows. Since reflectance and range images are fundamentally and precisely aligned, we first colorize the reflectance image by using its similarity with a color image, then transfer the color to the 3D geometric model.

The proposed technique does not require precise calibration between the color and range images. Such calibration is required for conventional texture mapping. Thus, the gaps or discontinuities in appearance due to calibration errors can be avoided even if several images must be registered on a model. Moreover, the appearance of the entire surface of a model can be assigned from a partial view of the model if the appearance does not change significantly or have a repetitive pattern in the entire model.

In the following sections, we introduce the details of the proposed technique: assigning color seed points in the reflectance image based on a local similarity between the color and reflectance images, and colorizing the reflectance image based on the seed points.

4.2.1 Image colorization

Image colorization is a technique for adding color to a monochrome image and has been used in some specific applications, such as coloring monochrome movies or creating color-coded images for electron photomicrograph or X-ray imaging. Since adding color values to a monochrome image has no clearly defined procedure, the current approaches attempt to estimate all colors based on clues given as seed points manually[59][60][61] or automatically[62][63][64].

Yatziv et al.[60] proposed a fast image colorization technique using Dijkstra’s distance[65]. This technique estimates the color at each pixel in a grayscale image by calculating a weighted average of Dijkstra's distances from each seed point which has chromatic information. Dijkstra's distance is computed considering changes in luminance in the monochrome image. If the change in luminance from the seed point is small, the chromatic information at the seed point is mainly copied to the pixel. More precisely, if the color is described in YCbCr color space, the color of each monochrome pixel is estimated using Dijkstra's distance as follows:

$$c_i = \frac{\sum_{j \in \Omega_c} w(i,j) c_j}{\sum_{j \in \Omega_c} w(i,j)} \tag{4.1}$$

$$w(i,j) = {r_{ij}}^{-\alpha} \tag{4.2}$$

$$r_{ij} = \min \sum_{k=1}^{n-1} |Y_{p_{k+1}} - Y_{p_k}|_{p_{k+1} \in N(p_k), p_1 = i, p_n = j} \tag{4.3}$$

where c_i is the estimated color value(CbCr) in pixel i, Ω_c is a set of seed points which have color information, r_{ij} is Dijkstra's distance from pixel i to pixel j, α is a gain parameter that controls the effect of weighting function $w(i,j)$ based on Dijkstra's distance, and Y_{p_k} and $N(p_k)$ are the intensity and the neighbor pixels of the pixel p_k.

Equation (4.2) indicates that the seed point, which has a small Dijkstra's distance to the target pixel, is preferentially selected to colorize the monochrome pixel in the grayscale image. In contrast, a seed point which has a Dijkstra's path with large luminance change contributes little to the color estimation of the monochrome pixel.

4.2.2 Colorization of range image using reflectance image

Based on the image colorization technique described above, we propose a new colorization technique for a 3D geometric model utilizing a reflectance image and a color image. The basic idea of the proposed technique is as follows; since reflectance and range images are fundamentally and precisely aligned, we colorize the reflectance image using its similarity with a color image at first, then transfer the color to the range image. In the following sections, we introduce our techniques, which determines correspondences in reflectance and color images using HOG features[66].

Assignment of seed points in a reflectance image

We colorize a reflectance image obtained by a laser scanner based on Yatziv's method [60] shown in Section 4.2.1. To colorize it, we first assign seed points that have chromatic information in the reflectance image. We adopt the following two approaches for assigning seed points.

1. Manual assignment by human intervention
2. Automatic assignment by determining the corresponding regions between reflectance and color images

In the manual assignment, a 3D geometric model is colorized according to the human instruction. Several seed points are selected manually in the reflectance image, and color information in the color image is assigned to these seed points.

In order to assign color seed points automatically, Simple Linear Iterative Clustering (SLIC) [67] and Histograms of Oriented Gradients (HOG) features [66] are utilized. SLIC proposed by Achanta et al. [67] is a technique to divide an image into small segments with similar size called “superpixels”. Each segment extracted by SLIC holds pixels that have similar intensity or color. HOG is proposed by Dalal et al. [66] for pedestrian detection in camera images. HOG is able to describe local object appearances robustly according to the distribution of gradient orientation of the intensity.

First, we divide the reflectance and color images into small segments with SLIC, and then the local features are extracted by applying HOG to small regions around the segments (Fig.4.1). Note that we apply Canny edge detection to the reflectance and color images to compare the outlines of objects for HOG feature extraction. Finally, based on the similarities of HOG features between the small regions in the reflectance

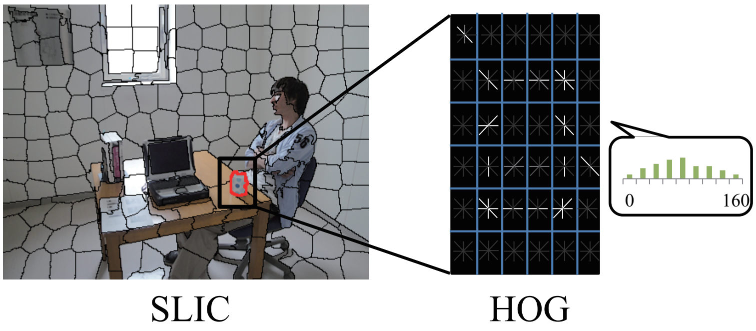


Figure 4.1: Assignment of seed points by local features with SLIC and HOG

and color images, the correspondences of the regions in both images are determined. Seed points in the reflectance image are chosen at the center of the segmented regions, and the chromatic information is copied from the center of the corresponding regions to the color image.

Here, in order to exclude mismatches between reflectance and color images as much as possible, we adopt 2-step corresponding determination technique: First, we find the most similar 5 corresponding points in a color image to each point in a reflectance image with a large HOG window to get global correspondence, and then select the suitable one from the 5 correspondences with a small HOG window which can describe the appearances in detail. Here, we consistently defined the window size as around 20% and 10% of input images for large and small HOG windows, respectively. We also eliminate correspondences whose feature similarities are too low.

Colorization of reflectance image using range image

Based on the color seed points given by the method described above, the reflectance image can be colorized with conventional colorization techniques used for monochrome images. These techniques are quite effective. However, it is sometimes difficult to colorize a reflectance image properly because the reflectance values indicate intensities on the surfaces of target objects under a single-frequency light source, and the object boundaries in the reflectance image are rather indistinct due to the lack of sufficient appearance information in comparison with the camera image.

In contrast, a range image includes the structural edges of target objects as jump edges that can be detected easily. Focusing on the fact that the range and reflectance images are precisely aligned, we can use the jump edges in the range image as additional edges for colorization of the reflectance image.

Consequently, we developed a new colorization technique extending Yatziv's method [60] so that edges not only in the reflectance image but also in the range image are considered simultaneously for colorization of the reflectance image.

To take these edges in the range image into account, we define a new energy function using a exponential function instead of Eq. (4.3) as follows (See A.1).

$$r_{ij} = \min \sum_{k=1}^{n-1} |Y_{p_{k+1}} - Y_{p_k}| e^{|D_{p_{k+1}} - D_{p_k}|}{}_{p_{k+1} \in N(p_k), p_1 = i, p_n = j} \tag{4.4}$$

where D_{p_k} is an intensity value at pixel p_k in a range image. Dijkstra's distance is computed by considering the changes in intensity in the monochrome reflectance image and the range image. If no clear edge exists along the path from a seed point to a target pixel in both the reflectance and the range images, Dijkstra's distance becomes small and the chromatic information at the seed point affects the color estimation of the target pixel significantly. In contrast, the seed point with large Dijkstra's distance due to jump edges in the reflectance and/or the range image slightly influences the color estimation.

Proposed method

The proposed colorization techniques for a 3D geometric model are summarized as follows. Note that the YCbCr color space is used in this research, however, the proposed technique works in other color spaces such as YUV or $l\alpha\beta$.

Method 1 : Manual assignment of seed points

1. Acquire range and reflectance values and a color image by a laser scanner and a digital camera, respectively.

2. Create range and reflectance images in which gray values of each pixel are proportional to the measured range and reflectance values.

3. Assign seed points in the reflectance image manually according to the correspondence between the reflectance and the color images.

4. Apply the proposed colorization technique in Eqs. (4.1), (4.2), and (4.4) using range and reflectance images, and obtain a colorized reflectance image. Luminance Y in the colorized reflectance image is determined by a reflectance value.

5. Transfer the color value of each pixel in the colorized reflectance image to the corresponding range image and construct a colorized 3D model from the range image.

Method 2 : Automatic assignment of seed points

To assign seed points automatically, we roughly determine the correspondence between reflectance and color images using HOG features. To do so, 3 in the method 1 is replaced as follows.

3. Divide reflectance and color images in small segments using SLIC, and determine the correspondence between segments according to HOG features. Then, assign the color information to the center pixel of each segment in the reflectance image from the corresponding region in the color image.

4.3 Experiments

4.3.1 Experiments with LIDAR

This section introduces the results of the colorization experiments. Range and reflectance images are obtained by the 3D laser measurement robot[4]. This robot captures surrounding range and reflectance data by rotating the laser scanner (SICK, LMS151) on a rotary table. Color images are taken by a digital camera (Fujifilm,

(a) Scene 1

(b) Scene 2

(c) Scene 3

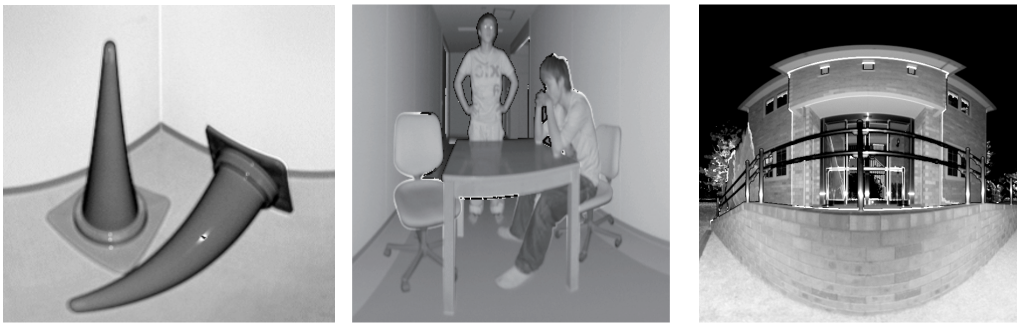
Figure 4.2: Experimental setup

FinePix S7000) by hand. In the experiments, the parameter α is set as $\alpha = 6$, and the number of superpixels is determined for each image experimentally so that we can get small super pixels enough to describe the appearance pattern on target objects in it. Range and reflectance images are quantized to 255 levels with the linear normalization (See A.2).

Figure 4.2 shows three experimental conditions: a simple environment with two road cones of different colors (scene 1), a complex environment with a human and other objects (a table and chairs) (scene 2), and a house made of red bricks (scene 3). Figure 4.3 shows the range and reflectance images of these scenes captured by the measurement robot.

First, we assigned seed points to the reflectance images in Fig.4.3(b) manually (Fig.4.4) and automatically (Fig.4.5). Note that only 10% correspondences are shown

(a) Range image(Scene 1) (c) Range image(Scene 2) (e) Range image(Scene 3)

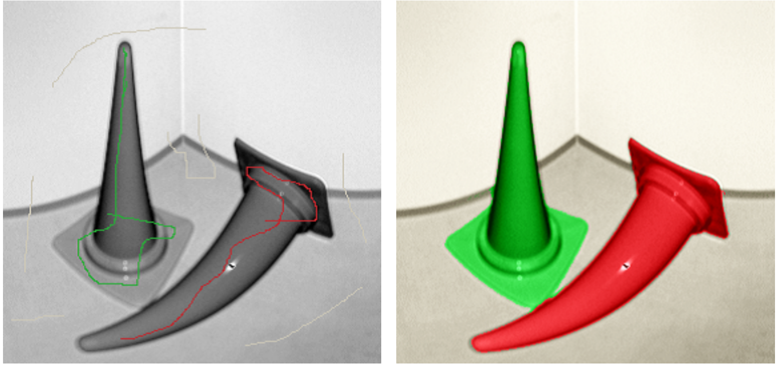


(b) Reflectance image(Scene 1) (d) Reflectance image(Scene 2) (f) Reflectance image(Scene 3)

Figure 4.3: Range and reflectance images in each scene

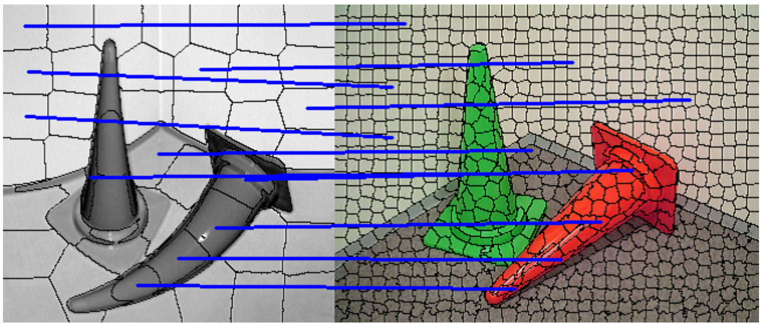
in Fig.4.5 for visibility. Figures 4.4(a) and 4.5(b) show the seed points in the reflectance image in both methods. Colorized reflectance images are shown in Fig.4.4(b) and Fig.4.5(c), respectively. Figure 4.6(a) is the 3D mesh model constructed from the range image shown in Fig.4.3(a). Figures 4.6(b)(c) show the 3D mesh model colorized by the method 1 (manual) and the method 2 (automatic), respectively. It is clear that the proposed methods successfully add color information to the surface of the 3D geometric models without accurate pose estimation.

Next, we carried out the experiments using a reflectance image and a color image taken in scene 2 (Fig.4.2(b)). Similar to the experiments in the scene 1, we assigned seed points manually (Fig.4.7) or automatically (Fig.4.8). Figures 4.7(a) and 4.8(b) show assigned seed points in both methods. Colorized reflectance images are shown in Fig.4.7(b) and Fig.4.8(c), respectively. We also colorized Fig.4.8(b) by applying Yatziv's method[60] which doesn't use range information, and the result is shown in

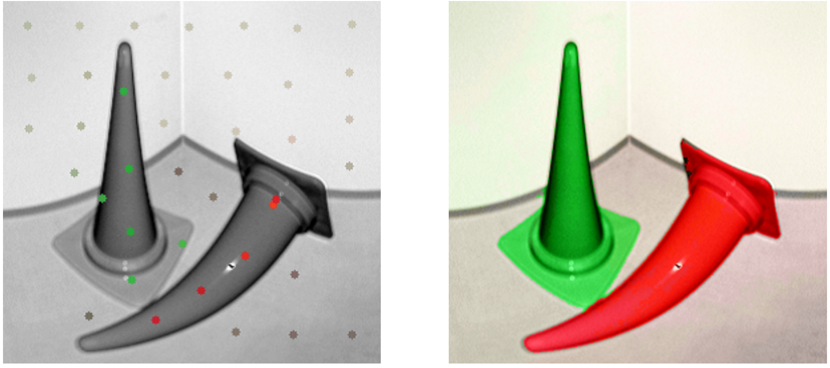


(a)Seed points assigned manually (b)Colorization result

Figure 4.4: Proposed method 1 in scene 1 (manual)



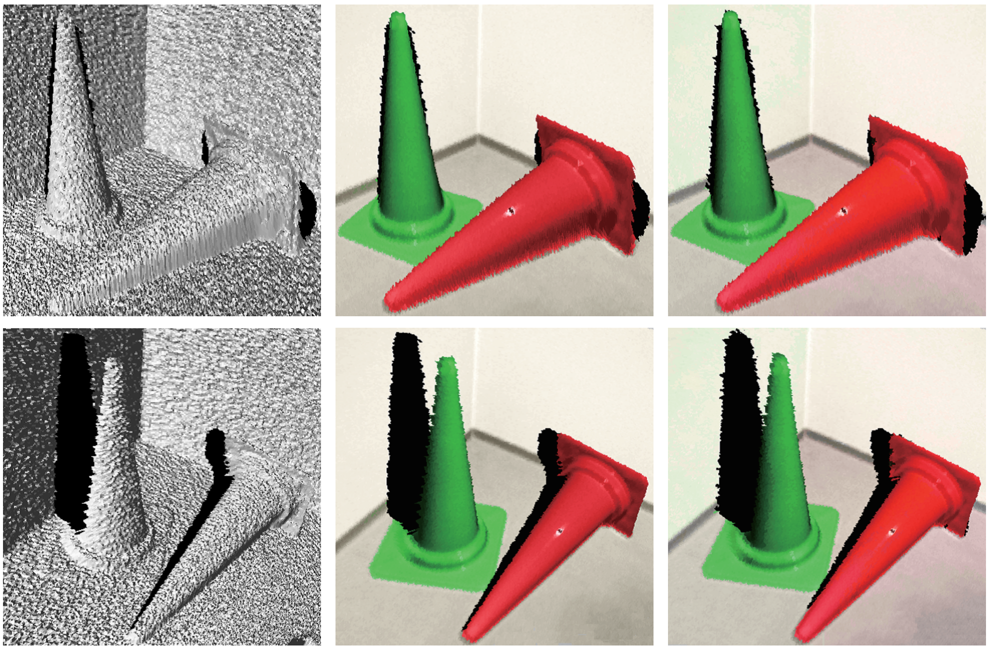
(a)The correspondence between Fig.4.2(a) and Fig.4.3(b)



(b)Seed points assigned automatically (c)Colorization result

Figure 4.5: Proposed method 2 in scene 1 (automatic)

Fig.4.8(d). Figures 4.8(c)(d) show that the proposed methods successfully prevent these color seeds from spreading extremely by considering jump edges in a corresponding range image. The colorized 3D model for scene 2 is shown in Fig.4.9. Fig-

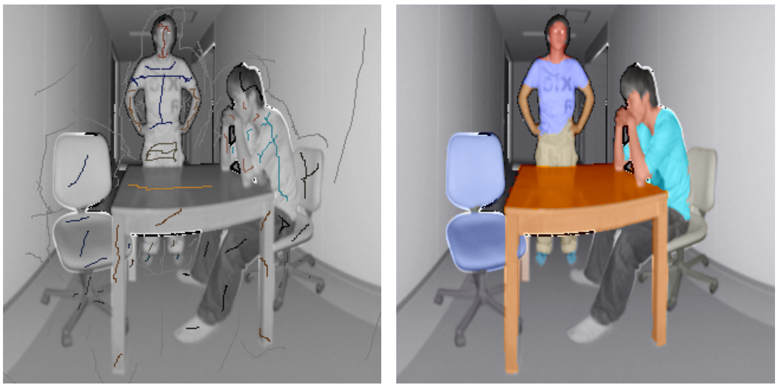


(a)Original 3D mesh model (b)Proposed method 1 (c)Proposed method 2

Figure 4.6: Colorized 3D geometric model in scene 1

ures 4.9(a)-(c) are the 3D mesh model constructed from the range image (Fig.4.3(c)), the 3D mesh model colorized by the method 1 (manual) , and the one colorized by the method 2 (automatic), respectively. From these results, we verified that the proposed methods are capable of creating colorized 3D geometric models in more complex scene.

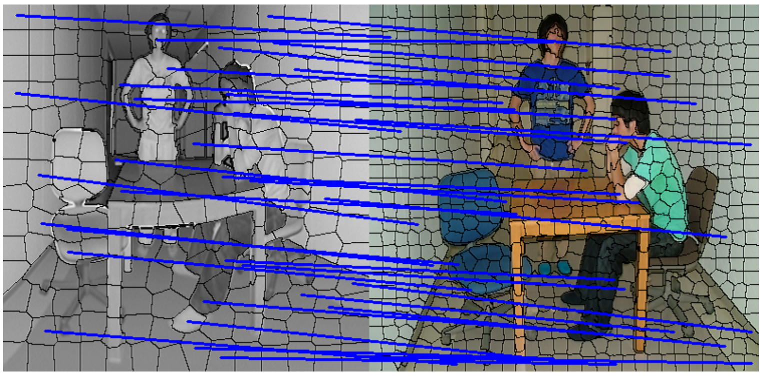
Finally, we carried out experiments in scene 3 (Fig.4.2(c)). In this experiment, we colorized an entire 3D geometric model of a house made of red bricks with a partial view of the house shown in Fig.4.2(c). The entire 3D geometric model is created from four range and reflectance images shown in Figs.4.3(e)(f). Figure 4.10 shows colorized reflectance images using the method 1 manually, and the method 2 automatically. The colorization results of the 3D geometric model (Fig.4.11(a)) in scene 3 are shown in Figs.4.11(b)(c). Note that while the manual assignment can give color seed points to the regions which don't appear in the picture if an operator assigns all the correspondences, the automatic assignment maps color seeds only to objects with



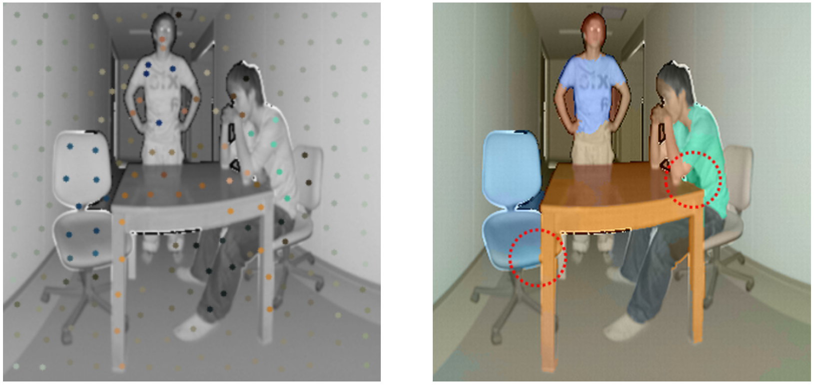
(a)Seed points assigned manually (b)Colorization result

Figure 4.7: Proposed method 1 in scene 2 (manual)

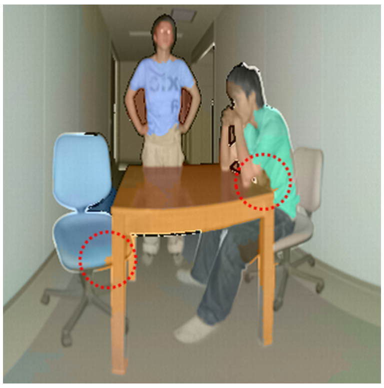
similar appearances in the picture, for instance, a house and trees. This causes colorless regions in Fig.4.10(c). To overcome this problem, we can use multiple pictures from different viewpoints and obtain additional color seed points by finding correspondences among reflectance images and the pictures. Figure 4.11(d) demonstrates the colorization result using three color images taken from different viewpoints with the proposed method 2, and you can see that the colorized regions increased. Consequently, we successfully colorized the entire geometric model using a partial view of the target house.



(a) The correspondences between Fig.4.2(b) and Fig.4.3(d)



(b) Seed points given automatically (c) Colorization result

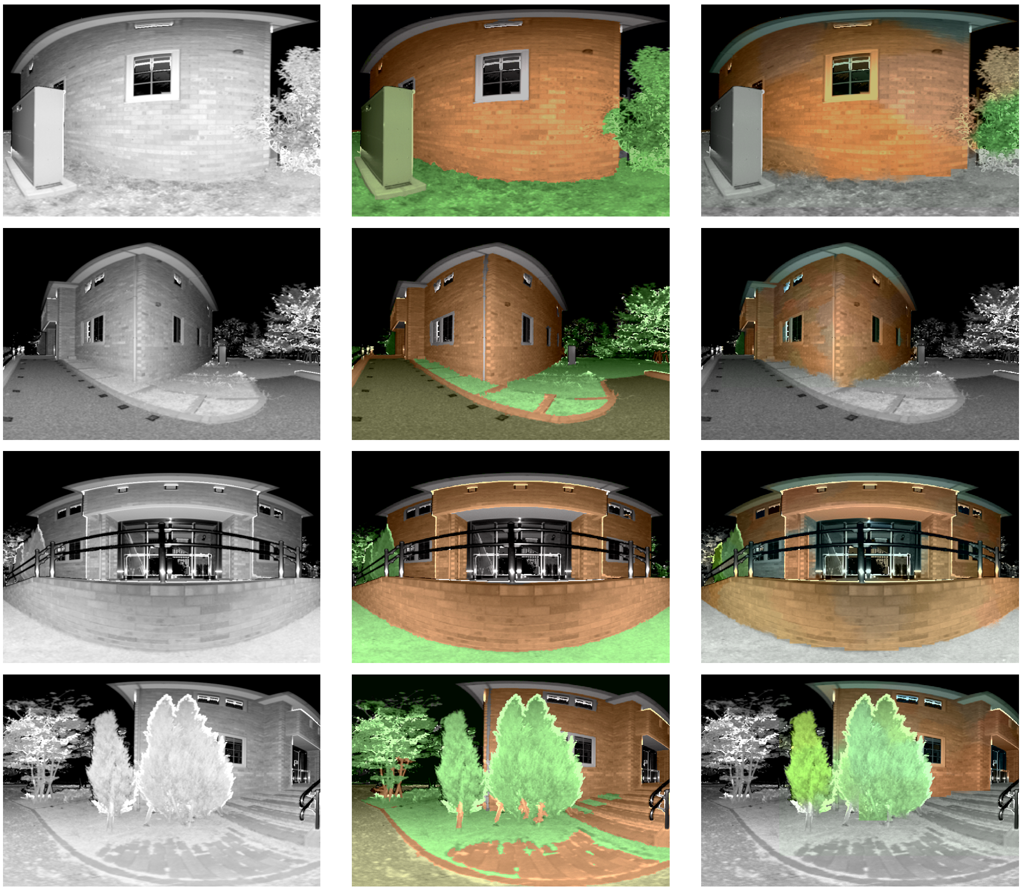


(d) Colorization result without range image[60]

Figure 4.8: Proposed method 2 in scene 2 (automatic)

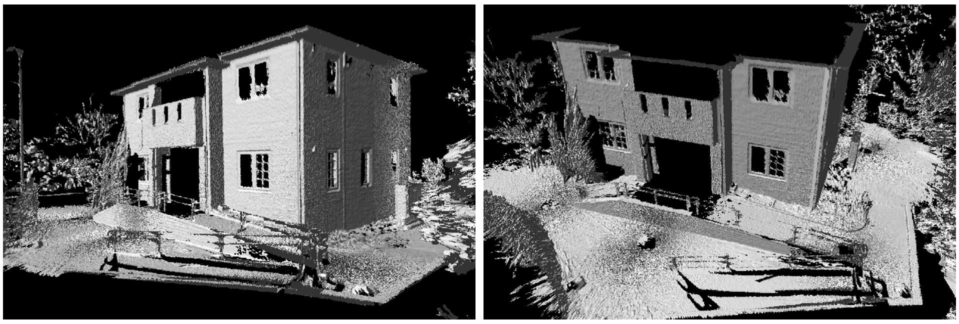
(a)Original 3D mesh model (b)Proposed method 1 (c)Proposed method 2

Figure 4.9: Colorized 3D geometric model in scene 2

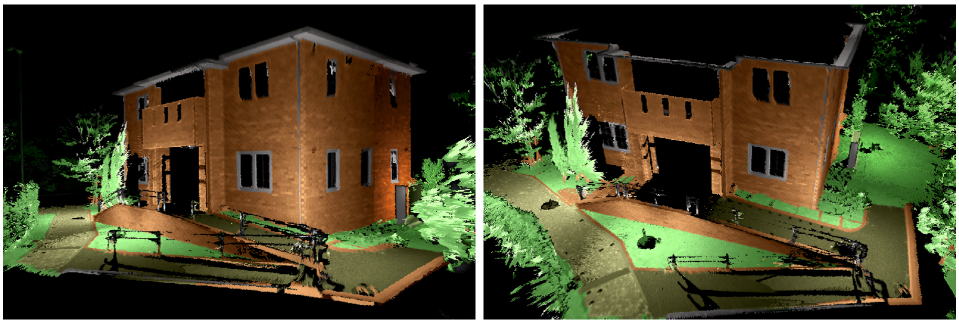


(a) (b) (c)

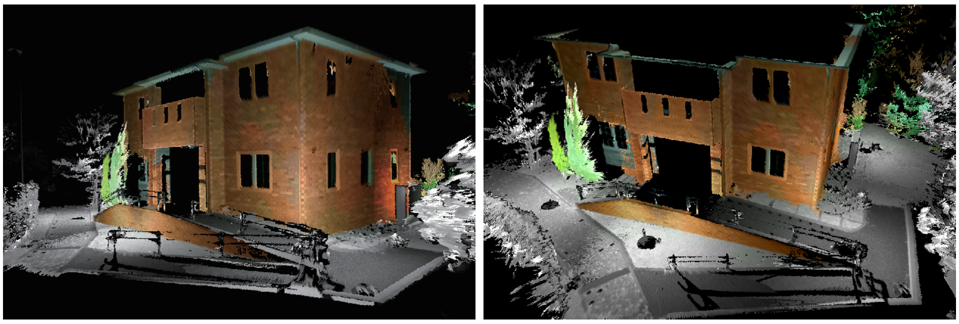
Figure 4.10: **Colorization of reflectance images from an only single picture.** (a) Original reflectance images. (b) Colorized reflectance images with proposed method 1 (manual). (c) Colorized reflectance images with proposed method 2 (automatic).



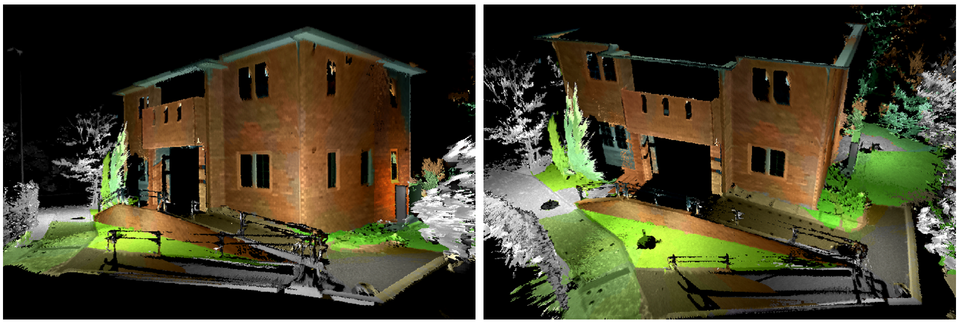
(a)Original 3D mesh model



(b)Proposed method 1 (manual)



(c)Proposed method 2 (automatic)



(d)Proposed method 2 using three pictures taken from difference viewpoints

Figure 4.11: Colorization of the entire 3D geometric model

4.3.2 Quantitative evaluation

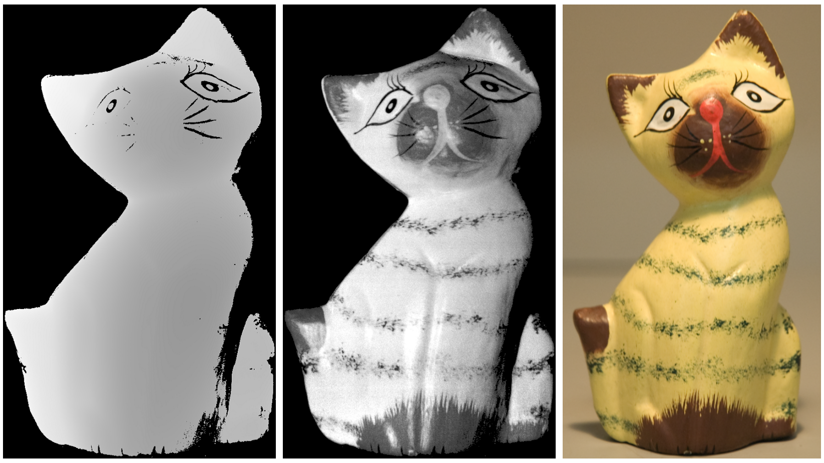
We demonstrates the validity of the proposed colorization technique for photo-realistic 3D modeling in comparison with texture mapping. In this experiments, range and reflectance images are acquired with the ShapeGrabber system (with scan head SG-100 on a PLM300 linear displacement mechanism) and FARO Focus3D (FARO) and as shown in Fig.4.12 and Fig.4.16, respectively.

We first assign several color seed points manually with the human instruction (proposed method 1, Fig.4.13(a) and Fig.4.17(a)), and automatically according to the correspondence between the reflectance image and the color image with SLIC and HOG features (proposed method 2, Fig.4.13(b)-(e) and Fig.4.17(b)(c)). Second, colorization of the reflectance images is performed based on the color seed points. The colorization results are shown in Fig.4.14(a)-(c) and Fig.4.18(a)-(c). Finally, the color information in the colorized reflectance images is transferred to each corresponding 3D point (Fig.4.15(a) and Fig.4.19(a)), and the colorized 3D models are constructed as shown in Fig.4.15(b)(c) and Fig.4.19(b)-(d).

In an experiment with the cat model shown in Fig.4.12, we colorized the entire model using the proposed methods. Note that we used only one texture image (Fig.4.12(c)), and colorized the back side of a target object by assigning seed points using a typical front side texture (Fig.4.13(d)(e)). These experimental results show that the proposed methods successfully add color information to the surface of the 3D geometric models without accurate pose estimation from only a partial view of the target object.

In another experiment with the building model shown in Fig.4.16, we demonstrate the validity of the proposed colorization method by comparing it with the Yatziv's method. Fig.4.18(a)(b) show the colorization results using the proposed method and the Yatziv's method based on color seeds in Fig.4.17(a). Since the proposed method can prevent color seeds from diffusing too much by detecting abrupt changes in geometry, the buildings are colorized appropriately based on the seed points even when reflectance values are quite similar between adjacent buildings. In contrast, the Yatziv's method allowed some color seeds to spread excessively and gave the colors over the adjacent buildings (Fig.4.18(b)). The proposed method 2 also colorized most of the building facades correctly, but some parts were assigned inappropriate colors due to mismatches between the reflectance image and texture image. This experiment is rather challenging since these images were taken from distant viewpoints and their appearances are not similar in some regions. However, this result shows that we need to improve the performance of the automatic color seed assignment even more.

Additionally, we created/measured texture-mapped 3D geometric models (Fig.4.15(d)[21] and Fig.4.19(e)) to compare the performances between the proposed method and texture mapping. In the comparison of Fig.4.15(b)(c) with Fig.4.15(d), and Fig.4.19(b)(c) with Fig.4.19(e), the qualities of 3D models constructed by the proposed methods are as good as those of texture mapping. We also carried out quantitative evaluations of the proposed methods. In YCbCr color space normalized 0 to 255, the differences of color values in the colorized reflectance images and the texture images at each pixel are also calculated in Cb and Cr channels, and the root mean square (RMS) errors between the texture image (Fig.4.15(d) and Fig.4.19(e)) and colorized reflectance images (Fig.4.15(b)(c) and Fig.4.19(b)-(d)) are shown in Table 4.1 and Table 4.2. Note that YCbCr space represents color as brightness and two different color signals, and the RMS errors in Cb and Cr channels are affected slightly by the change of brightness, such as the cast shadow in Fig.4.12(c). Since the RMS errors are within 3.12% and 5.35% of the color ranges of the texture images respectively, the proposed methods are as useful as conventional texture mapping techniques for creating photo-realistic 3D models.



(a) Range image (b) Reflectance image (c) Texture image

Figure 4.12: Range, reflectance, and texture images of a target cat model

Table 4.1: RMS errors in Cb and Cr color space (Cat model)

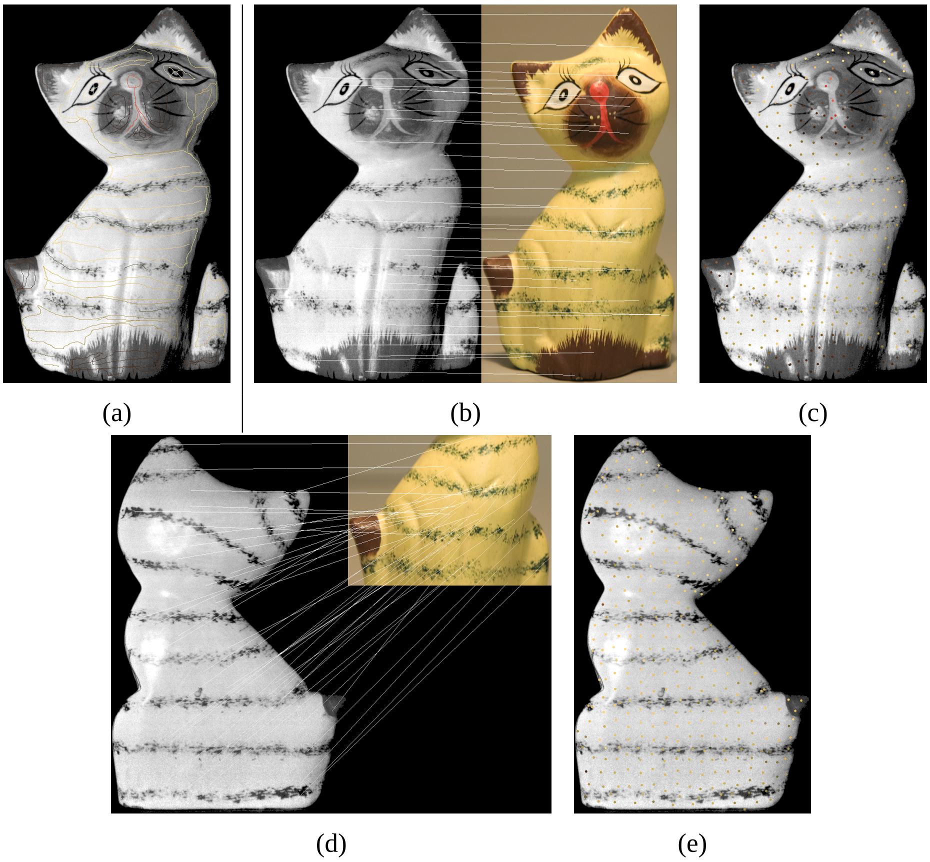
Cb and Cr ranges are 113-214 and 50-126 in Fig.4.16(c). RMS errors between the texture image (Fig.4.15(d)) and colorized reflectance images (Fig.4.15(b)(c)) are computed by calculating the differences of color values in the texture images and colorized reflectance images at each pixel.

		RMS error		Error percentage	
		Cb	Cr	Cb	Cr
Proposed method 1	front(Fig.4.15(b):upper)	2.57	2.68	2.54%	3.12%
	back(Fig.4.15(b):lower)	2.16	2.46	2.14%	2.86%
Proposed method 2	front(Fig.4.15(c):upper)	2.28	2.15	2.26%	2.50%
	back(Fig.4.15(c):lower)	2.26	2.60	2.24%	3.02%

Table 4.2: RMS errors in Cb and Cr color space (Building model)

Cb and Cr ranges are 109-148 and 110-162 in Fig.4.19(d). RMS errors between the texture images (Fig.4.19(d)) and colorized reflectance images (Fig.4.18(a)-(c)) are computed by calculating the differences of color values in the texture images and colorized reflectance images at each pixel.

	RMS error		Error percentage	
	Cb	Cr	Cb	Cr
Proposed method 1 based on Fig.4.17(a) (Fig.4.18(a))	1.79	2.78	4.59%	5.35%
Yatziv's method based on Fig.4.17(a) (Fig.4.18(b))	1.80	2.78	4.61%	5.35%
Proposed method 2 based on Fig.4.17(c) (Fig.4.18(c))	2.06	2.74	5.28%	5.27%


(a) (b) (c)

(d) (e)

Figure 4.13: **Seed point assignment to the cat model.** (a) Seed points applied manually (Proposed method 1). (b) The correspondences between Fig.4.12(b) and Fig.4.12(c). (c) Seed points applied automatically based on the correspondences (Proposed method 2). (d) Seed point assignment for the back side using a typical front side texture. We can assign proper seed points even when the back side texture images are not available. (e) Seed points applied automatically using a typical pattern texture (Proposed method 2).

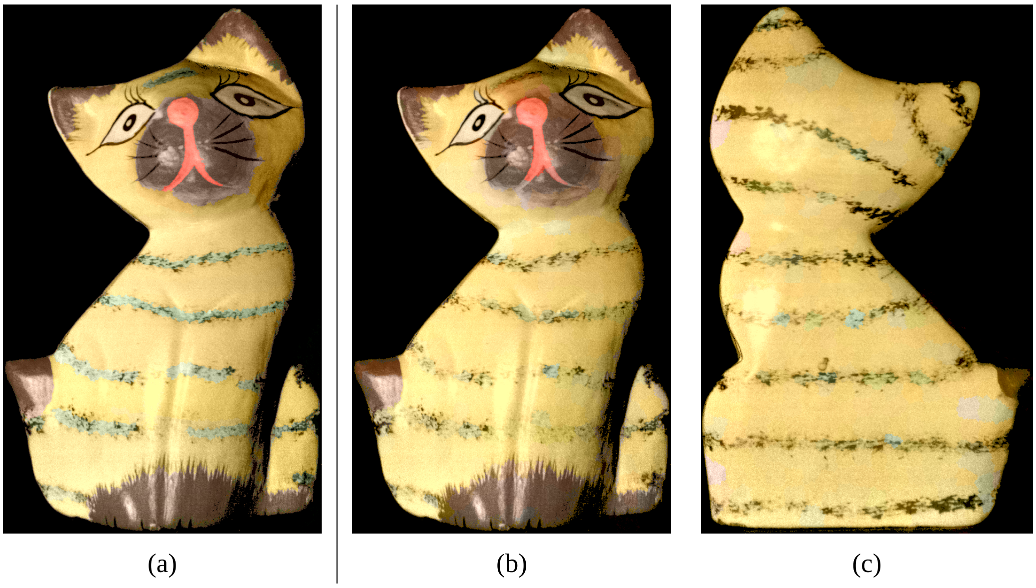
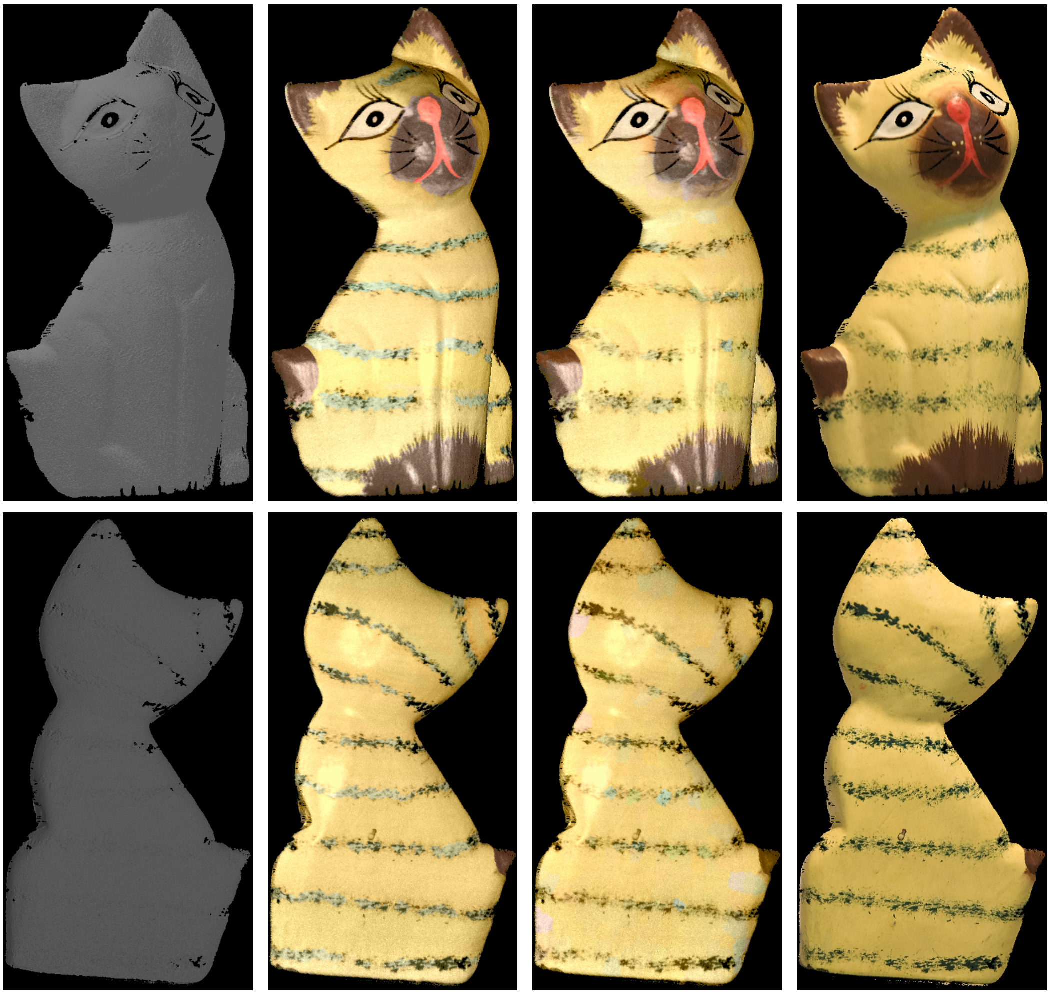

(a) (b) (c)

Figure 4.14: **Colorization results of the cat model.** (a), (b) and (c) are colorized reflectance images based on Fig.4.13(a) (Proposed method 1) and Figs.4.13(c) and (e) (Proposed method 2), respectively.

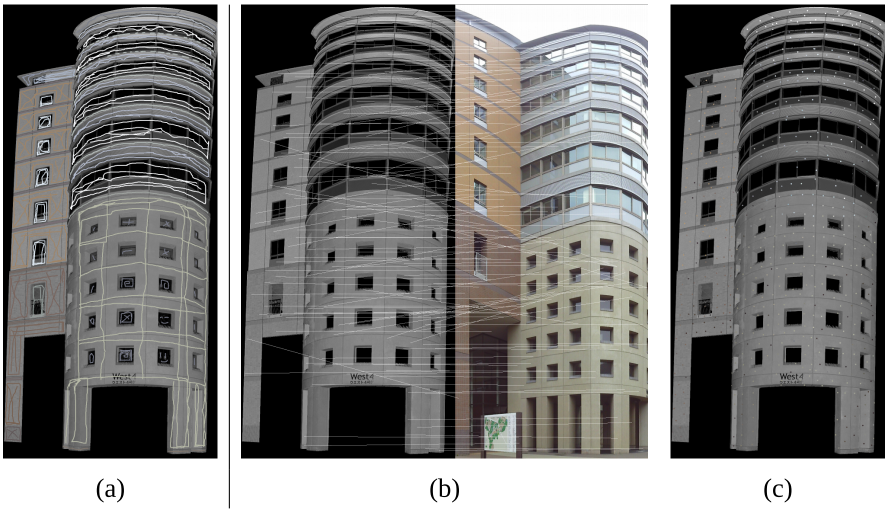


(a) Original 3D model (b) Method 1 with only Fig.4.12(c) (c) Method 2 with only Fig.4.12(c) (d) Texture mapping with multiple images

Figure 4.15: 3D models constructed by the proposed methods and texture mapping with the cat model

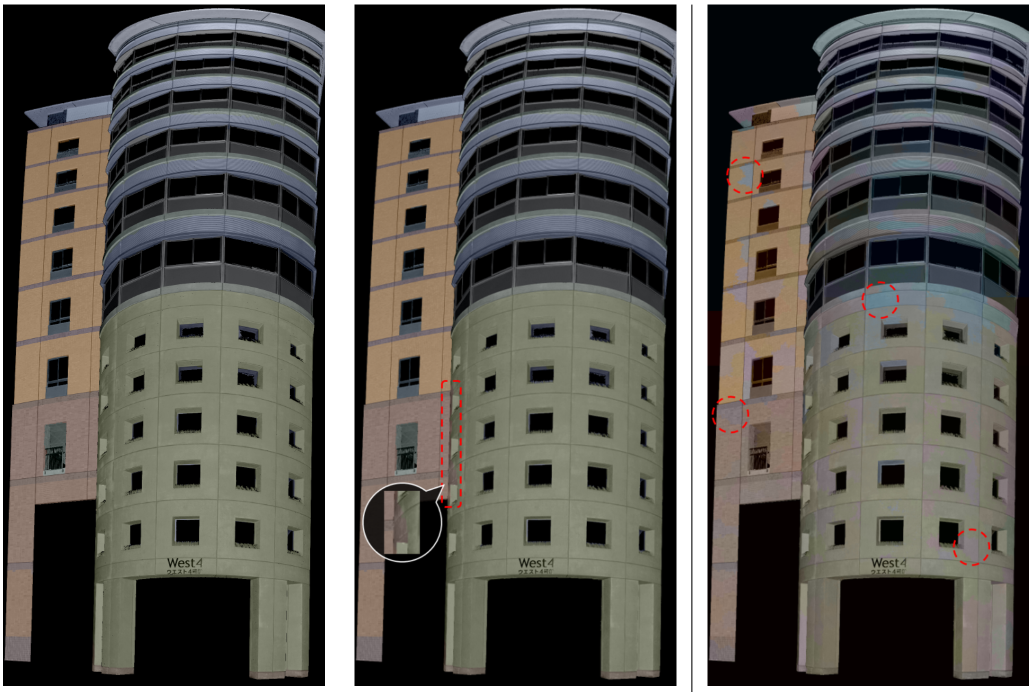
(a) Range image (b) Reflectance image (c) Texture image

Figure 4.16: Range, reflectance, and texture images of target buildings



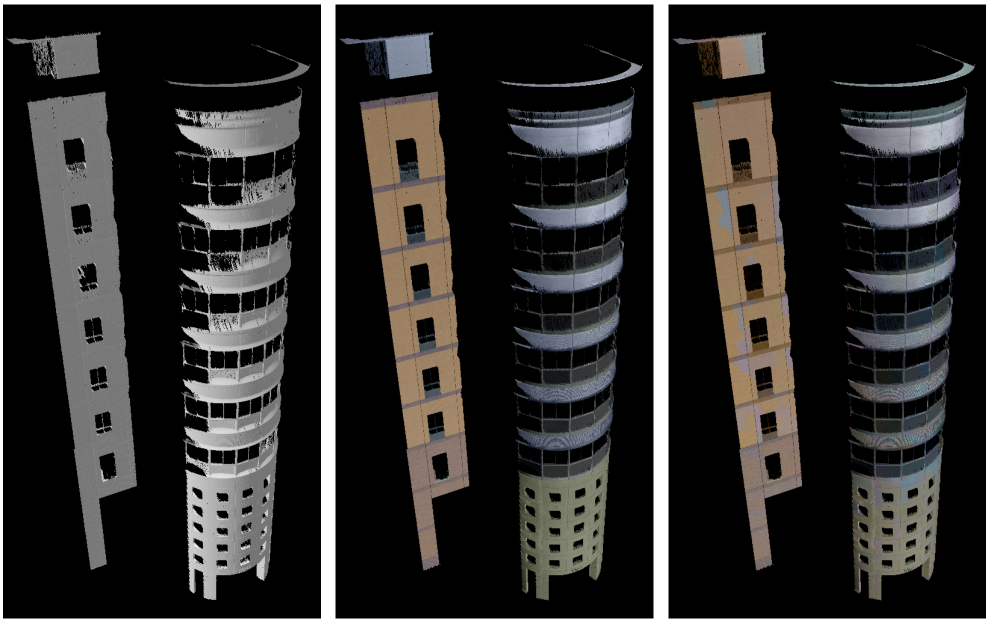
(a) (b) (c)

Figure 4.17: **Seed point assignment to the building model.** (a) Seed points applied manually (Proposed method 1). (b) The correspondences between Fig.4.16(b) and Fig.4.16(c). (c) Seed points applied automatically based on the correspondences (Proposed method 2).

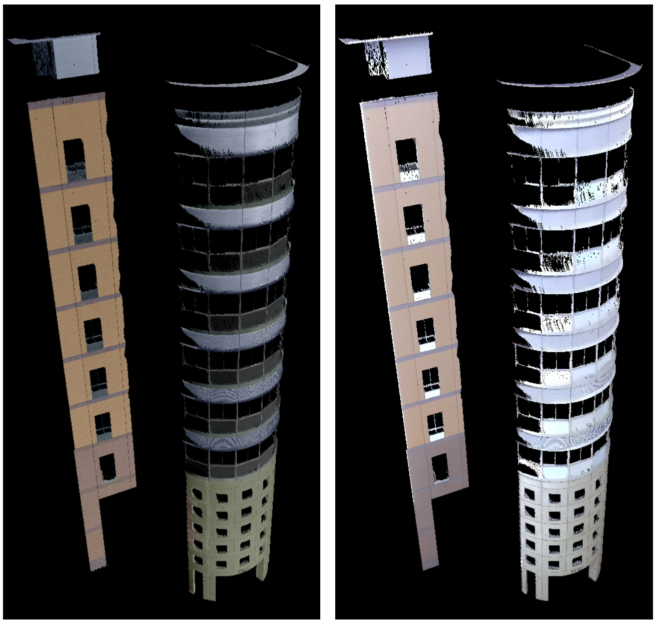


(a) (b) (c)

Figure 4.18: **Colorization results of the building model.** (a),(b) Colorized reflectance image based on Fig.4.17(a) with Proposed method 1 and Yatziv's method respectively. (c) Colorized reflectance images based on Fig.4.17(c) with Proposed method 2.



(a) Original 3D model (b) Method 1 (c) Method 2



(d) Yatziv's method (e) Texture mapping

Figure 4.19: 3D models constructed by the proposed methods and texture mapping with the building model

4.4 Conclusion

In this chapter, we proposed a new colorization technique for a 3D geometric model using the laser reflectivity. The proposed technique adds color information on the surface of a 3D geometric model by colorizing the reflectance image manually or automatically. Automatic colorization is easy but needs adequate correspondence between the reflectance and color images. In contrast, manual colorization gives a colored model even if the correspondence is slight between the images.

Precise calibration between color and range images is not required, even though it is indispensable for conventional texture mapping. Thus, gaps or discontinuities in appearance can be avoided even if several images must be registered on a model. Moreover, the appearance of the entire surface of a model can be assigned from a partial view of the model, if the appearance does not change significantly in the entire model.

In addition, we carried out a quantitative evaluation of the proposed colorization techniques in comparison with texture mapping to demonstrate their validities. The results show that the proposed techniques are able to create photo-realistic 3D models by colorizing reflectance images based on several color seed points, and the qualities of 3D models constructed by the proposed techniques are as good as those of texture mapping.

5

Texture Synthesis for Hole-Free Texture Mapping

In recent years, 3D geometric modeling of real-world objects with laser scanners has become very popular and widely used in many areas, such as robotics, landscape surveying, digital archive of cultural heritages[4], and so on. Once a 3D geometric model of a target object is obtained, a more realistic virtual model can be created by adding color information to the surface of the geometric model. For photo-realistic modeling, texture mapping is an effective technique and still remains one of important research topics in Virtual Reality (VR). In texture mapping, appearance information is added to a 3D geometric model by mapping texture images on the surface. However, in case that a 3D geometric model and a texture image are acquired from different viewpoints through two independent sensors such as a laser scanner and a digital camera, some texture-less regions (holes) may exist on the object surface where the appearance information is missing due to the occlusion by obstacles or out-of-sight of the camera (Fig.5.1).

In this chapter, we propose a novel hole-free texturing technique utilizing laser reflectivity. The laser reflectivity, which denotes the power of a reflected laser light/pulse, is obtained as by-product of the range information at laser scanning. By aligning the reflectivity according to the order of scanning, a reflectance image, which contains rich appearance information of the target surface, is obtained. This image is quite similar to a camera image since each pixel indicates the intensity on the surface of the targets under a single-frequency light source. Owing to this similarity, we first determine a completion order of a texture image in an occluded region according to the reflectance value, and then synthesize the partial texture by copying and pasting a proper texture patch from the other regions with similar color and laser reflectivity.

The rest of this chapter is organized as follows. In Section 5.1, we provide an overview of the previous texture completion approaches. In Section 5.2, we describe the texture completion technique, including the generation of database patches, the definition of completion priority based on the laser reflectivity, and the completion algorithm. In Section 5.3, we show some experimental results to verify the effectiveness of the proposed method. In Section 5.4, we conclude this chapter and mention the future work.

5.1 Related work

For hole-free texture mapping, it is necessary to copy or synthesize a proper texture image in the holes from the other regions or other camera images. Boehm[68] proposed a simple texturing method by capturing multiple texture images from different

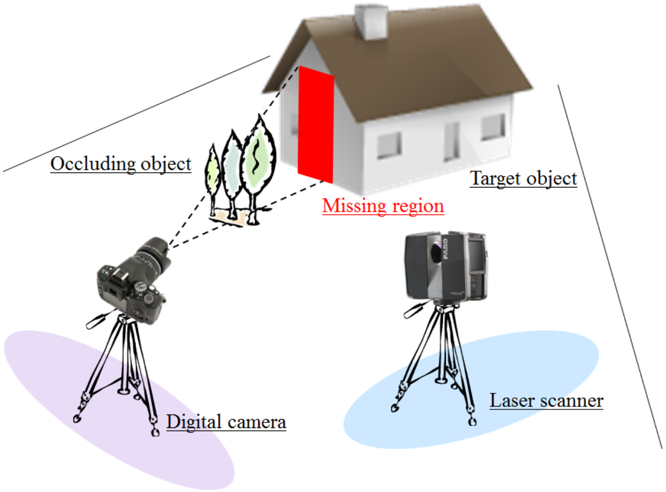


Figure 5.1: In case that the positions of the camera and the laser scanner differ from each other, some texture-less regions (holes) may exist on the object surface where the appearance information is missing due to the occlusion or out-of-sight of the camera.

viewpoints to suppress occluded regions. Ortin et al.[69] proposed a similar technique employing the information of adjacent images. These methods, however, reconstruct a virtual texture from multiple images taken from different viewpoints by assuming a planar target object and calculating a homography matrix between images. Moreover, the quality of the synthesized texture depends on how proper images are taken from different points of view.

On the other hand, Kawai et al.[70] proposed a completion method that restores the shape and the color of the object with geometrical and photometrical holes simultaneously. They selected proper regions with geometrical or photometrical information from other parts of the object by minimizing energy functions based on the similarities of geometries and textures. However, if the resolution of the shape and the color differ, blurry artifacts are appeared since this technique is a point-based approach.

Meanwhile, texture (or shape) inpainting are quite effective techniques to solve this problem. Komodakis et al.[71] proposed a patch-exemplar based method that solves texture inpainting as a discrete global optimization problem utilizing a belief propagation to avoid undesirable artifacts. For proper inpainting, the order of the completion

(a) Scene 1 (b) Scene 2

Figure 5.2: Input textures

is also quite important. Criminisi et al.[72], Sun et al.[73], and Li et al.[74] developed image completion techniques that determined the order of the completion by extending some curves or lines from a known region to an unknown region manually or automatically. Although these methods offered some excellent results, unfavorable discontinuities are still remained due to the geometrical and photometrical complexity of the target.

5.2 Texture completion utilizing laser reflectivity

In this section, we propose a new texture completion technique for a partially-occluded texture image, which extends patch-based completion techniques[71][72][73][74] to a reflectance image. In this research, we assume that the texture image is aligned to the geometrical model precisely beforehand[21]. Our first step is generating a database of small image patches from non-occluded regions in the texture image. We define the region of occluding objects Ω manually so that the rest of the texture image is specified as a source region Φ. Next, we determine the completion order (priority) of occluded region Ω taking into account structures in a reflectance image. Finally, we calculate the similarity between example patches in the database and the query patch centered at the top-prioritized pixel in Ω, and complete the texture image copying the most similar database patch to the missing region.

5.2.1 Generation of database patches

At the beginning of the proposed technique, we divide a input texture image I into a missing region Ω and a source region $\Phi(=\bar{\Omega})$ manually(Fig.5.2(a)(b)), where the former is supposed to have only reflectivity and the latter is supposed to have reflectivity and texture. Next, we create a example patch consisting of $k \times k$ pixels centered at a pixel in Φ, and construct a database D for all pixels in Φ. Note that we discard patches including pixels in the missing region. Moreover, each patch holds not only partial texture image but also the corresponding reflectance image, and both information describe the characteristic of each example patch.

5.2.2 Completion priority based on laser reflectivity

A noteworthy fact of image inpainting is that the completion order significantly affects the quality of the restored image[71][72][73][74]. Inspired by these researches, we define ”Priority map”, shown in Fig.5.3(b)(f) utilizing reflectivity. Since reflectance images have appearance information of the target objects, the structures of the scene can be extracted by detecting edges in the reflectance image. By applying Canny filter to the reflectance image, the priority is calculated as follows:

$$Priority(m) = \left\{ \frac{1}{k^2} \sum_{p \in N(m)} M(p) \right\} + C(m) \tag{5.1}$$

$$M(p) = \begin{cases} 1 & (p \in \Phi) \\ 0 & (otherwise) \end{cases}$$

Where, m denotes a pixel in the missing region Ω, $N(m)$ denotes a patch with $k \times k$ pixels centered at pixel m, and $C(m)$ is a value of pixel m in the canny-filtered reflectance image. According to $M(p)$, the pixel m on the boundary between the source region Φ and the missing region Ω has high priority. Furthermore, when the pixel m is located on a structural edge in the canny-filtered reflectance image, the priority becomes higher.

5.2.3 Completion algorithm

We complete the deteriorated texture image by copying the most similar example patch to the missing region. The similarity between a query patch q centered at the top-prioritized pixel in Ω and one of database patches p is calculated based on texture and reflectivity as follows:

$$Similarity(p,q) = \sum_{p \in D} \left\{ SSD(T_p, T_q) + \alpha SSD(R_p, R_q) \right\}^{-1} \tag{5.2}$$

$$SSD(I,J) = \sum_{u=0}^{k} \sum_{v=0}^{k} \{I(u,v) - J(u,v)\}^2 \tag{5.3}$$

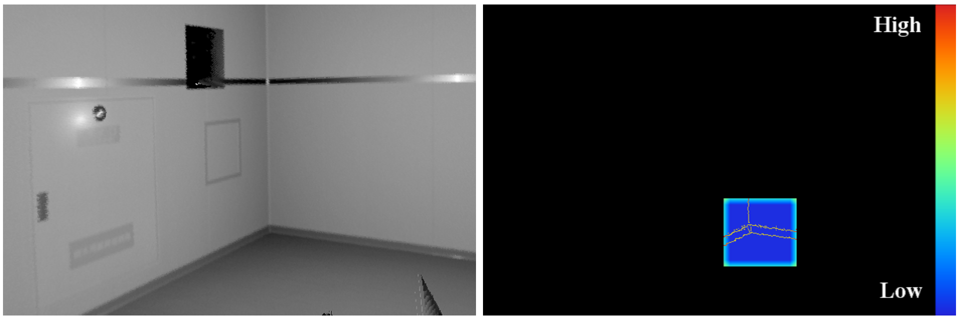
Where, T_p and R_p denote the texture and reflectance image belonging to patch p, and α is a gain parameter. Note that we synthesize a reflectance image taken from the viewpoint of the texture image by interpolating the reflectivity on 3D points (Fig.5.3(a)(e)). The best patch b for the query patch q is determined by finding the patch in D that maximizes the similarity, and complete the missing region in q by copying the texture image of b.

5.3 Experiments

This section introduces the results of texture completion experiments. Range and reflectance images are captured by a laser scanner (SICK, LMS151) and a turn table. The size of these images is 760×1135 pixels. Texture images are taken by a digital camera (Nikon, D300) and registered on the corresponding 3D geometric models[21]. We set parameters as $k = 15$, $\alpha = 1.0$ in the first experiment, and $k = 13$, $\alpha = 0.7$ in the second experiment.

We first carried out the preliminary experiment in a simple situation shown in Fig.5.2(a). In this experiment, we cut out a part of the input texture image and attempted to restore the missing region. Fig.5.3(a)(b) show a reflectance image and a priority map. Fig.5.3(c)(d) show the completion results without and with the reflectance image. By taking into account a property of the reflectance image, our proposed technique successfully filled in the missing region compared with the conventional ”onion-peeling” technique which merely fills the hole from outside to inside. Textured 3D models of Scene 1 are shown in Fig.5.4(a)-(c).

Next, we addressed the experiment in a more complex scene shown in Fig.5.2(b) where an obstacle occludes the wall and the texture behind it cannot be seen. Fig.5.3(e) shows the reflectance image from the same viewpoint of the texture image, and the completion order is determined by considering the structures of the reflectance image as shown in Fig.5.3(f). The completion results after applying the conventional technique and the proposed technique are shown in Fig.5.3(g)(h) respectively. 3D geometric models of Scene 2 after applying texture mapping are shown in Fig.5.4(d)-(f). As shown these figures, our proposed technique outperformed the conventional technique which caused some artifacts. However, the proposed technique also produced partial discontinuous textures, especially in a vertical direction. The reason why these artifacts occurred was that sufficient appearance information in the reflectance image could not be acquired due to the low resolution of the laser scanner.



(a) Reflectance image (b) Priority map

(c) Completed texture without reflectance image (d) Completed texture with reflectance image

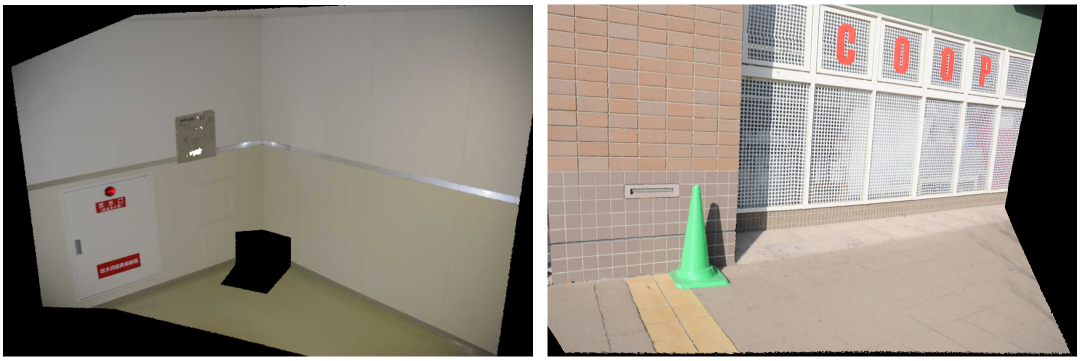
Experimental result (Scene 1)

(e) Reflectance image (f) Priority map

(g) Completed texture without reflectance image (h) Completed texture with reflectance image

Experimental result (Scene 2)

Figure 5.3: Texture completion results



(a) Original texture

(d) Original texture

(b) Completed texture without reflectance image

(e) Completed texture without reflectance image

(c) Completed texture with reflectance image

(f) Completed texture with reflectance image

Scene 1

scene 2

Figure 5.4: Texture mapping results

5.4 Conclusion

In this chapter, we proposed a new texture completion technique for texture mapping utilizing laser reflectivity. The proposed technique is able to generate the texture image in an occluded region of the 3D geometric model accurately by prioritizing the completion order.

In the future, we will improve the proposed technique so that it works well even with a low resolution reflectance image. Furthermore, we will extend our proposed method and generate the entire texture of a 3D geometric model from a single picture.

6

Conclusion and Outlook

Measuring 3D geometric information in the real world is one of the most essential tasks in research fields such as robotics, computer vision, and computer graphics. Its fundamental applications would include building 3D environmental maps available for mobile robot localization or creating 3D models for cultural heritages of famous buildings and art works.

In this dissertation, we first discussed the literature of 3D vision introducing several range measurement techniques such as stereo triangulation and time-of-flight ranging. We also focused on the characteristics of time-of-flight laser scanners in detail, and developed a set of algorithms for enhancing functionality of 3D modeling with a laser scanner employing the benefits of the laser reflectivity. This work actually addressed four different tasks: Range image smoothing using trilateral filter (Chapter 2), Range image completion with belief propagation (Chapter 3), Manual/Automatic colorization for 3D geometric models (Chapter 4), and Texture synthesis for hole-free texture mapping (Chapter 5). This chapter review these approaches and their contributions, and discuss outlook for the future work.

6.1 Range image smoothing using trilateral filter

The proposed smoothing filter ”Trilateral filter” allows us to successfully refine deteriorated range images including noise due to measurement error while preserving the geometric features. Although the proposed technique needs geometric and appearance information of the target object, it relies only on data measured with a laser scanner and no additional measurements or instruments are required. In addition, calibration error between the geometric and appearance data can be avoided that is a crucial problem for conventional appearance-based denoising techniques. Simulations using synthesized data and experiments using actual scan data demonstrated the validity of the proposed denoising technique and the performance in edge preservation. As future work, we should discuss the optimum parameters of the proposed technique. Besides, we currently assume that discontinuities in reflectance images are incidental to the geometric changes such as jump and roof edges, texture patterns sometimes affects the smoothing and completion results. In the future, we will develop a technique that can detects only geometric edges for better denoising results.

6.2 Range image completion with belief propagation

In order to restore holes in range images, we developed a new completion technique which estimates missing range data in optimization framework with belief propagation. Reflectance images tend to be more sensitive to geometric changes such as jump and roof edges than range images, and much edges can be observed more clearly in reflectance images. Taking advantage of the fact, we proposed 2-step completion technique: First restore a deteriorated reflectance image with belief propagation in order to acquire a rough estimate of an appearance of the target object, and then apply an extended belief propagation to a corresponding deteriorated range image which weights belief messages among adjacent pixels according to continuities of reflectance data in the restored reflectance image. While simulations and experiments verified that the proposed completion technique successfully restored missing regions in range images, it still depends on the parameters adjusted manually. As future work, we will improve the proposed method so that it can determines optimum parameters automatically according to input range data. We will also develop the completion technique further so that it can restore larger holes and more complicated structures.

6.3 Manual/Automatic colorization for 3D geometric models

While a number of studies has addressed 2D-3D alignment techniques for texture mapping in order to assign colors to 3D geometric models, we developed a novel approach taking advantage of laser reflectivity. Inspired by image colorization, the proposed method gives colors to a 3D model by colorizing the corresponding reflectance image directly. This novel approach allows us to easily create colorized 3D models based on manually or automatically assigned color seeds, and also to avoid color shift due to alignment error between the 3D model and camera image. Colorization experiments verified that the proposed technique appropriately assigned colors to the target 3D models, but we need to improve automatic color seed assignment for better colorization results. In addition, although quantitative evaluations demonstrated that the chromatic difference of colorized models and texture-mapped models were quite small, 3D models created with texture mapping have much more realistic appearances. This would be partly because that the proposed method uses raw laser reflectivity as the brightness of the target object, but it only indicates an intensity on the surfaces under a single-frequency light source. Therefore, as future work, we have to develop a method that corrects reflectance values so that the reflectance image has similar appearance to a monochrome camera image.

6.4 Texture synthesis for hole-free texture mapping

For removing obstacles or filling holes in texture images used for texture mapping, we proposed a novel texture synthesis technique based on laser reflectivity. A color image and a 3D model are usually captured from different viewpoints with different sensors such as a laser scanner and a digital camera, and some texture-less regions (holes) may exist on the object surface where the appearance information is missing due to the occlusion or out-of-sight of the camera. In order to complete the missing regions avoiding artifacts, the proposed method first extracts structural information in the missing region from corresponding reflectance image and determines the completion order, and then synthesizes texture according to the completion order. This approach allows us to inpaint undesirable parts in the input texture images, but still has some issues to be solved. The size of texture patch affects completion results, however, we determine it experimentally for each input image at the moment. So as to generate appropriate size patches automatically, we will develop a technique that adjust the patch size according to texture frequency around the missing region. Additionally, we will introduce optimization process to avoid artifacts due to discontinuities of assigned patches.

A

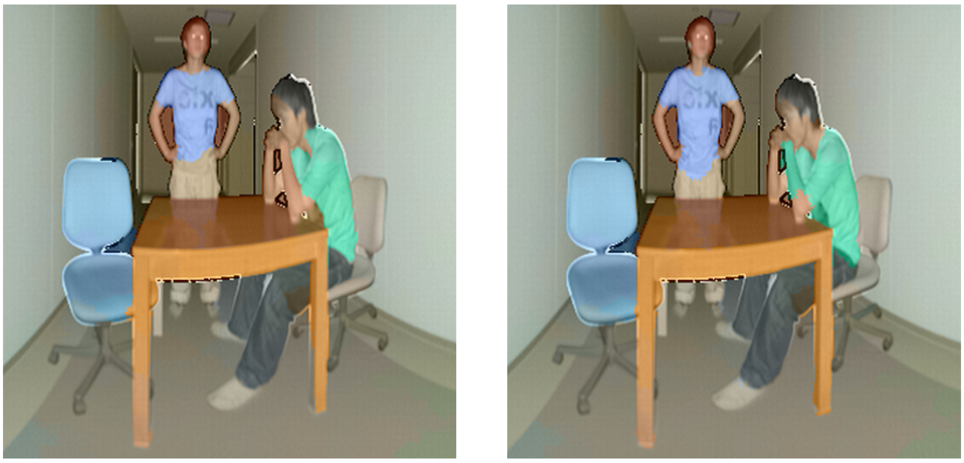
Additional Discussion of Chapter 4

A.1 Cost function for colorization

Color information from a seed point should be diffused within a single object, and should not be spread over different objects. To do so, we emphasize a gap in a depth image rather than a reflectance image and use an exponential cost function for a depth image. To demonstrate the validity of the exponential cost function, we also carried out an experiment with the cost function weighted by the difference of reflectance values instead of Equation (4.4). The formulation is as follows:

$$r_{ij} = \min \sum_{k=1}^{n-1} \left\{ \kappa |D_{p_{k+1}} - D_{p_k}| + 1 \right\} |Y_{p_{k+1}} - Y_{p_k}| \tag{A.1}$$

where κ denotes a gain parameter to control the effect of the depth data for color estimation. You can see the colorization results using Yatziv ’ s method in Fig. 4.8(d), exponential weight in Fig. 4.8(c), and difference weight in Fig. A.1. As shown in Fig. 4.8(d), Yatziv ’ s method which considers only reflectance data couldn’t prevent some color seeds from spreading too much, especially an abdominal region of the person sitting on a chair. We tried to colorize this region appropriately with Equation (A.1) changing the parameter κ (Fig. A.1(a) and (b)). When κ is small, the colorization result was almost the same as that of Yatziv ’ s method. On the other hand, although bigger κ enabled us to colorize the abdominal region appropriately, the cost term of depth data became more dominant than that of reflectance data and some color seeds diffused too much ignoring edges in a reflectance image. This can be seen in an left arm of the person on a chair. In contrast, the exponential cost function could successfully detect abrupt changes in geometry, which are object boundaries, and prevent undesirable color diffusions over different objects (Fig. 4.8(c)).



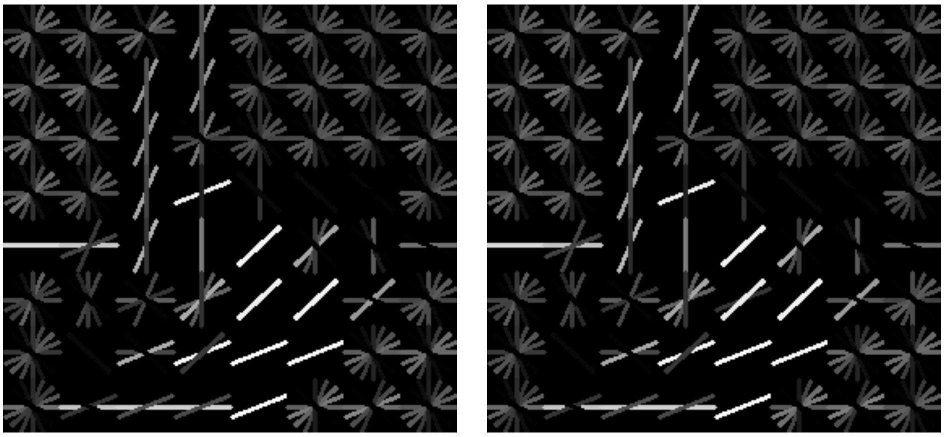
(a) Colorization result with small κ ($\kappa = 1$) (b) Colorization result with large κ ($\kappa = 100$)

Figure A.1: Colorization with a cost function weighted by a difference of reflectance using Fig. 4.8(b)

A.2 HOG feature from raw/normalized reflectivity

Although reflectance data is quantized to 256 levels, HOG feature still can be extracted stably. In order to extract HOG feature, the gradient orientation calculated at each pixel in an input image is discretized in large intervals, and the histogram at each cell is computed robustly by casting a weighted vote for the corresponding bin according to the gradient and strength of each pixel within the cell.

To demonstrate it, we extracted HOG features with 8 bins (45 degree/per a bin) and 8×8 cells from raw and normalized range images individually and compared the difference. As shown in Fig. A.2, the extracted HOG features changed slightly.



(a) (b)

Figure A.2: **HOG extraction based on reflectance data in scene 1.** (a) and (b) visualize extracted HOG features from normalized and raw reflectance values, respectively.

B

Sensor Specifications

This appendix shows a specification of each range sensor that was used for experiments in every chapter. Note that LMS200, LMS151, LMS511 (SICK), and Focus3D (FARO) are time-of-flight laser scanners while SG-100 (ShapeGrabber) is an active stereo sensor which was used for a quantitative evaluation in Chapter 4.

Sensor	Measuring range	Accuracy	Field of view	Angular resolution
 LMS200	80 [m]	± 15 [mm]	180 [deg.]	0.25 [deg.]
 LMS151	50 [m]	± 30 [mm]	270 [deg.]	0.25 [deg.]
 LMS511	80 [m]	± 25-50 [mm] (1...30[m])	190 [deg.]	0.167 [deg.]
 Focus3D	120 [m]	± 2 [mm]	305 [deg.]	0.009 [deg.]
 SG-100	190 [mm]	70-125 [μm]	90-190 [mm]	70-150 [μm/point]

Table B.1: Laser scanner specifications

Bibliography

[1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):pp.105–112, October 2011.

[2] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, 29(1):141–159, 2010.

[3] G.J. Iddan and G. Yahav. 3d imaging in the studio (and elsewhere...). *In Proceedings of SPIE*, 4298:pp.48–56, 2001.

[4] Ryo Kurazume, Yusuke Noda, Yukihiro Tobata, Kai Lingemann, Yumi Iwashita, and Tsutomu Hasegawa. Laser-based geometric modeling using cooperative multiple mobile robots. In *Proc. IEEE International Conference on Robotics and Automation*, pages 3200–3205, 2009.

[5] Roger N. Clark. *Visual Astronomy of the Deep Sky*. Cambridge University Press and Sky Publishing, 1990.

[6] Andrew J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the Ninth IEEE International Conference on*

Computer Vision - Volume 2, ICCV ’03, pages pp.1403–. IEEE Computer Society, 2003.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.

[8] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, nov 2004.

[9] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, jun 2008.

[10] Stephan Leutenegger, Margarita Chli, and Roland Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.

[11] M. Okutomi and Takeo Kanade. A multiple-baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.

[12] Andrea Fusiello, Emanuele Trucco, and Alessandro Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision Application*, 12(1):16–22, jul 2000.

[13] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’07)*, Nara, Japan, November 2007.

[14] M. A. Cazorla D. Viejo, J. M. Saez and F. Escolano. Active stereo based compact mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 529–534, 2005.

[15] Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki. Dense 3d reconstruction from high frame-rate video using a static grid pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(9):pp.1733–1747, 2014.

[16] C. L. Zitnick S. B. Kang, J. A. Webb and T. Kanade. A multi baseline stereo system with active illumination and real-time image acquisition. In *International Conference on Computer Vision (ICCV)*, pages 88–93, 1995.

[17] J. P. McDonald C. W. Urquhart, J. P. Siebert and R. J. Fryer. Active animate stereo vision. In *British Machine Vision Conference (BMVC)*, pages 75–84, 1993.

[18] Keiji Nagatani, Yoshito Okada, Naoki Tokunaga, Seiga Kiribayashi, Kazuya Yoshida, Kazunori Ohno, Eijiro Takeuchi, Satoshi Tadokoro, Hidehisa Akiyama, Itsuki Noda, Tomoaki Yoshida, and Eiji Koyanagi. Multirobot exploration for search and rescue missions: A report on map building in robocuprescue 2009. *Journal of Field Robotics*, 28(3):373–387, 2011.

[19] Megumi Shinozaki, Masato Kusanagi, Kazunori Umeda, Guy Godin, and Marc Rioux. Correction of color information of a 3D model using a range intensity image. *Computer Vision and Image Understanding*, 113:1170–1179, 2009.

[20] Ryo Kurazume, Ko Noshino, Zhengyou Zhang, and Katushi Ikeuchi. Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute. In *Proc. of Fifth Asian Conference on Computer Vision (ACCV)*, pages 99–106, 2002.

[21] Ryo Inomata, Kenji Terabayashi, Kazunori Umeda, and Guy Godin. Registration of 3d geometric model and color images using sift and range intensity images. In *Proceedings of the 7th international conference on Advances in visual computing - Volume Part I*, ISVC'11, pages 325–336. Springer-Verlag, 2011.

[22] Gabriel Taubin. A signal processing approach to fair surface design. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 351–358, New York, NY, USA, 1995. ACM.

[23] Mathieu Desbrun, Mark Meyer, Peter Schro”der, and Alan H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *SIGGRAPH ’99*, pages 317–324, 1999.

[24] Thouis R. Jones, Frédo Durand, and Mathieu Desbrun. Non-iterative, feature-preserving mesh smoothing. In *SIGGRAPH ’03: ACM SIGGRAPH 2003 Papers*, pages 943–949, New York, NY, USA, 2003. ACM.

[25] Shachar Fleishman, Iddo Drori, and Daniel Cohen-Or. Bilateral mesh denoising. *ACM Transactions on Graphics*, 22(3):950–953, 2003.

[26] A. Miropolsky and A. Fischer. Reconstruction with 3D geometric bilateral filter. In *SM ’04: Proceedings of the ninth ACM symposium on Solid modeling and applications*, pages 225–229, Aire-la-Ville, Switzerland, Switzerland, 2004. Eurographics Association.

[27] J. Diebel and S. Thrun. An application of Markov random fields to range sensing. In *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2005. MIT Press.

[28] M.Bohme, M.Haker, T.Martinetz, and E.Barth. Shading constraint improves accuracy of time-of-flight measurements. *Computer Vision and Image Understanding*, 114:1329–1335, 2010.

[29] James Davis Diego Nehab, Szymon Rusinkiewicz and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *ACM Transactions on Graphics(Proc. of ACM SIGGRAPH 2005)*, volume 24(3), pages 536–543, 2005.

[30] Takayuki Okatani and Koichiro Deguchi. Optimal integration of photometric and geometric surface measurements using inaccurate reflectance/illumination knowledge. In *CVPR2012*, 2012.

[31] Rolf Adelsberger, Remo Ziegler, Marc Levoy, and Markus Gross. Spatially adaptive photographic flash. In *ETH Zurich, Department of Computer Science, Technical Report No. 612,* 2008.

[32] Derek Chan, Hylke Buisman, Christian Theobalt, and Sebastian Thrun. A noise-aware filter for real-time depth upsamplingw. In *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.

[33] Marvin Lindner, Martin Lambers, and Andreas Kolb. Sub-pixel data fusion and edge-enhanced distance refinement for 2D/3D images. *International Journal on Intelligent Systems Technology and Application*, 5(3/4):344–354, 2008.

[34] R. Crabb, C. Tracey, A. Puranik, and J. Davis. Real-time foreground segmentation via range and color imaging. In *In Proc. of CVPR Workshop on Time-of-flight Computer Vision,* pages 1–5, 2008.

[35] Johannes Kopf, Michael Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007),* 26(3):96, 2007.

[36] Jean-Denis Durou, Maurizio Falcone, and Manuela Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Comput. Vis. Image Underst.*, 109(1):22–43, 2008.

[37] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV ’98: Proceedings of the Sixth International Conference on Computer Vision,* pages 839–846, Washington, DC, USA, 1998. IEEE Computer Society.

[38] Shin Yoshizawa, Alexander Belyaev, and Hans peter Seidel. Smoothing by example: Mesh denoising by averaging with similarity based weights. In *In Proceedings of the IEEE International Conference on Shape Modeling and Applications (2006),* pages 38–44. IEEE, 2006.

[39] Durand F. Eisemann E. Flash photography enhancement via intrinsic relighting. In *In ACM Transactions on Graphics (Proc. SIGGRAPH 04)*, pages 673–678, 2004.

[40] Petschnigg G., Szeliski R., Agrawala M., Cohen M., Hoppe H., and Toyama K. Digital photography with flash and no-flash image pairs. In *In ACM Transactions on Graphics (Proc. SIGGRAPH 04)*, pages 664–672, 2004.

[41] Mcmillan L. Benett E. P. Video enhancement using per-pixel virtual exposures. In *In ACM Transactions on Graphics (Proc. SIGGRAPH 05) (2005)*, pages 845–852, 2005.

[42] Markus Weber, Marco Milch, Karol Myszkowski, Kirill Dmitriev, Przemyslaw Rokita, and Hans-Peter Seidel. Spatio-temporal photon density estimation using bilateral filtering. In *Computer Graphics International 2004 (CGI'04)*, pages 120–127, 2004.

[43] Norihiko Kawai, Tomokazu Sato, and Naokazu Yokoya. Surface completion by minimizing energy based on similarity of shape. In *Proc. IEEE Int. Conf. on Image Processing (ICIP2008)*, pages 1532–1535, 2008.

[44] Jacob Becker, Charles Stewart, and Richard J.Radke. Lidar inpainting from a single image. In *In Proceedings of the IEEE International Workshop on 3-D Digital Imaging and Modeling*, 2009.

[45] S.Xu, A.Georghiades, H.Rushmeier, J.Dorsey, and L.McMillan. Image guided geometry inference. In *Proc. 3rd Int. Symp. on 3DPVT*, pages 310–317, 2006.

[46] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1), 2006.

[47] N. Komodakis and G. Tziritas. Image completion using global optimization. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 442–452, 2006.

[48] J.F. Blinn and M.E. Newell. Texture and reflection in computer generated images. In *Communications of the ACM 19 (10)*, pages 542–547, 1976.

[49] K. Yoshida and H. Saito. Registration of range image using texture of high-resolution color images. In *IAPR Workshop on Machine Vision Applications*, pages pp.150–153, 2002.

[50] Peter J. Neugebauer and Konrad Klein. Texturing 3d models of real world objects from multiple unregistered photographic views. In *In Eurographics '99 Conference proceedings*, pages 245–256, 1999.

[51] P. Viola and W.M.Well. Alignment by maximization of mutual information. *IJCV*, Vol.24(No.2):pp.137–154, 1997.

[52] Ioannis Stamos and Peter K. Allen. Integration of range and image sensing for photorealistic 3d modeling. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, pages pp.1435–1440, 2000.

[53] Yuya Iwakiri and Toyohisa Kaneko. Pc-based realtime texture painting on real world objects. In *In Eurographics 2001 Conference proceedings*, volume vol. 20, pages 105–113, 2001.

[54] Hendrik P.A. Lensch, Wolfgang Heidrich, and Hans Peter Seidel. Automated texture registration and stitching for real world models. In *Proceedings of the 8th Pacific Conference on Computer Graphics and Applications*, PG '00, pages 317–. IEEE Computer Society, 2000.

[55] Christophe Dumont Faysal Boughorbal, David L. Page and Mongi A. Abidi. Registration and integration of multi-sensor data for photo-realistic scene reconstruction. In *In: Proc. Applied Imagery Pattern Recognition*, pages 74–84, 1999.

[56] Kazunori Umeda, Guy Godin, and Marc Rioux. Registration of range and color images using gradient constraints and range intensity images. In *Proceedings of*

the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03, ICPR '04, pages 12–15. IEEE Computer Society, 2004.

[57] Becker S Boehm J. Automatic marker-free registration of terrestrial laser scans using reflectance features. In *In: 8th Conf. on Optical 3D Measurement Techniques*, pages 325–336, 2007.

[58] Shuji Oishi, Ryo Kurazume, Yumi Iwashita, and Tsutomu Hasegawa. Colorization of 3d geometric model utilizing laser reflectivity. In *2013 IEEE International Conference on Robotics and Automation (ICRA2013)*, 2013.

[59] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *SIGGRAPH2004*, 2004.

[60] Liron Yatziv and Guillermo Sapiro. Fast image and video colorization using chrominance blending. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 15(5), 2006.

[61] P. Lagodzinski and B. Smolka. Colorization of medical images. In *APSIPA ASC 2009*, 2009.

[62] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to grayscale images. In *ACM Siggraph Conference*, pages pp.277–280, 2002.

[63] Y. Tai, J. Jia, and C. Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume Vol. 1, 2005.

[64] T. Horiuchi. Colorization algorithm for gray-level image by probabilistic relxation. In *Proceedings of IEEE International Conference on Pattern Recognition*, 2003.

[65] E.Dijkstra. A note on two problems in connexion with graphs. In *In Numerische Mathematik*, 1959.

[66] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition*, volume Vol.1, 2005.

[67] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[68] J. Boehm. Multi-image fusion for occlusion-free facade texturing. In *IAPRS Vol. XXXV, part 5*, pages 867–872, 2004.

[69] D. Ortin and F. Remondino. Generation of occlusion-free images for realistic texture mapping. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, page 36(5/W17), 2005.

[70] N. Kawai, A. Zakhor, T. Sato, and N. Yokoya. Surface completion of shape and texture based on energy minimization. In *Proc. International Conference on Image Processing*, pages 913–916, 2011.

[71] G. Tziritas N. Komodakis. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. In *IEEE Trans. Image Process. 16 (11)*, pages 2649–2661, 2007.

[72] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based inpainting. In *IEEE Trans. Image Process. 13 (9)*, pages 1200–1212, 2004.

[73] J. Jia J. Sun, L. Yuan. Image completion with structure propagation. In *ACM Trans. Graphics 24 (3)*, pages 861–868, 2005.

[74] Ming Zhao Shutao Li. Image inpainting with salient structure completion and texture propagation. In *Pattern Recognition Letters 32 (9)*, pages 1256–1266, 2011.