Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project

A. Banno · T. Masuda · T. Oishi · K. Ikeuchi

Abstract We have been conducting a project to digitize the Bayon temple, located at the center of Angkor-Thom in the kingdom of Cambodia. This is a huge structure, more than 150 meters long on all sides and up to 45 meters high. Digitizing such a large-scale object in fine detail requires developing new types of sensors for obtaining data of various kinds related to irregular positions such as the very high parts of the structure occluded from the ground. In this article, we present a sensing system with a moving platform, referred to as the Flying Laser Range Sensor (FLRS), for obtaining data related to these high structures from above them. The FLRS, suspended beneath a balloon, can be maneuvered freely in the sky and can measure structures invisible from the ground. The obtained data, however, has some distortion due to the movement of the sensor during the scanning process. In order to remedy this issue, we have developed several new rectification algorithms for the FLRS. One method is an extension of the 3D alignment algorithm to estimate not only rotation and translation but also motion parameters. This algorithm compares range data of overlapping regions from ground-based sensors and our FLRS. Another method accurately estimates the FLRS's position by combining range data and image sequences from a video camera mounted on the FLRS. We evaluate these algorithms using a IS-based method and verify that both methods achieve much higher accuracy than previous methods.

Keywords Large-scale 3D scenes · Cultural sites · Moving range sensors · Alignment · Shape rectification

A. Banno (✉) · T. Masuda
Graduate School of Information Science and Technology, University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan
e-mail: vanno@cvl.iis.u-tokyo.ac.jp

T. Masuda
e-mail: tom@cvl.iis.u-tokyo.ac.jp

T. Oishi · K. Ikeuchi
Graduate School of Interdisciplinary Information Studies, University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan

T. Oishi
e-mail: oishi@cvl.iis.u-tokyo.ac.jp

K. Ikeuchi
e-mail: ki@cvl.iis.u-tokyo.ac.jp

1 Introduction

Preserving cultural heritage objects is a very important mission for us because they are deteriorating or being destroyed. Digital preservation, measuring such objects in three dimensions and representing them in digital forms, is one of the best ways to accomplish this preservation.

For the past seven years, we have been conducting a project entitled "the Great Buddha Project (Miyazaki et al. 2005; Ikeuchi et al. 2003)" to develop sensors and software for digital preservation of large-scale outdoor structures of cultural interest. Along with this project, we have also archived several cultural heritage objects such as the Kamakura Great Buddha in Kamakura, Japan, and the Atchana Buddha in Skhotai, Thailand. We are currently extending these projects by attempting the challenging project of digitizing the Bayon temple in the Angkor ruin in Cambodia.

The Bayon Temple (Fig. 1) is located at the center of Angkor-Thom in the Kingdom of Cambodia and unites the outlook and traditions of ancient India and the Khmer. The temple, constructed around the end of the 12th century, is well known for its size (more than 150 meters long and

up to 45 meters high), its 51 towers, its 173 calm, smiling faces carved on the towers, and its double corridors carved in beautiful and vivid bas-relief.

We have been digitizing the Bayon Temple with the co-operation of the Japanese government team for Safeguarding Angkor (JSA) since 2002. The motivation for this digitization project is twofold: the historical value of digitally preserving the Bayon Temple in the current form and the technical challenge of applying current technologies to large-scale cultural heritage objects. Regarding the technical side, we have encountered various issues and derived solutions to meet three goals:

1. To extend the geometric pipeline to determine the shape of extremely large-scale objects
2. To extend the photometric pipeline to be able to texture the site in the real color against strong and variable environmental illuminations
3. To develop new types of sensors that can measure regions invisible from conventional ground-based laser range sensors

For the first goal, alignment, which determines relative relations between range data sets, is one of the main components of our geometric pipeline. Traditional alignment algorithms (Besl and McKay 1992; Chen and Medion 1992; Zhang 1994) cannot be applied to our numerous range data


Fig. 1 The Bayon Temple
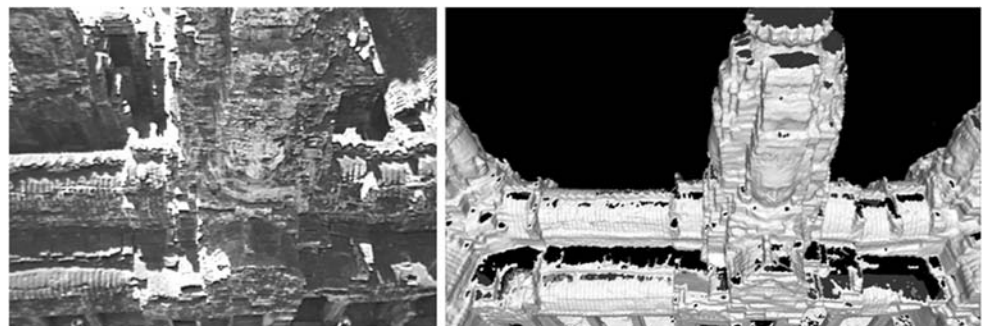
sets. We developed two kinds of alignment algorithms for large-scale site modeling. One is a rapid alignment algorithm using graphics hardware (Oishi et al. 2005). The other is an accurate simultaneous registration algorithm running on a PC cluster reducing the influence of error accumulation (Nishino and Ikeuchi 2002; Oishi et al. 2003).

With respect to the second goal, two issues, illumination and size, need to be considered in handling a large outdoor object such as the Bayon Temple. Compensation of illumination effects is necessary (Kawakami et al. 2005) because illumination conditions vary during the picture-taking session. The second issue is the size: aligning more than 1000 color images over one single geometric model needs a global bundle adjustment (Kurazume et al. 2002).

With respect to the third goal, digitizing large-scale structures requires developing new types of sensors for obtaining range data of various kinds situated in irregular positions. One class of such irregular positions is the very high portions of the structures occluded from the ground. Figure 2, for instance, shows a part of the 3D model taken by conventional laser range sensors, such as commercial sensors, which require a stable base. There are several regions without range data at the roofs, which are invisible from the ground. For scanning high positions, temporary scaffolds are often built around structures. However, this scaffold method is impractical for large cultural heritage objects because they detract from the appearance of these priceless objects and also shaking occurs at the top of the high scaffolds.

To overcome this difficulty, we proposed a novel 3D measurement system, a Flying Laser Range Sensor (FLRS) (Hirota et al. 2004), suspended from the underside of a balloon platform (Fig. 3). Several helicopter-based sensors (Miller and Amidi 1998; Thrun et al. 2003) have been proposed for large-scale site modeling. We avoided using a helicopter as the platform for the sensor for three reasons. First, the high frequency vibrations from the helicopter engine make it difficult to design rectification algorithms. Second, even though helicopters have high maneuverability, it is still not entirely safe to use them in proximity to very important cultural heritage structures. Third, helicopters are considered as strategic, and we cannot import and export them freely. Thus, our sensor was designed with a balloon platform.

Fig. 2 Lack of data for regions invisible from the ground

Fig. 3 The Flying Laser Range
Sensor
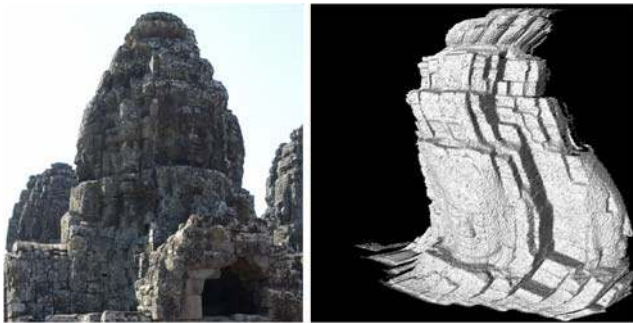
Fig. 4 A sample snapshot compared with the distorted range data obtained by the FLRS

Measurement by the FLRS has a unique problem: the FLRS moves during its scanning and thus provides distorted range data, as shown in Fig. 4. There are several works dealing with shape rectification and alignment of non-rigid objects (Feldmar and Ayache 1996; Szeliski and Lavallée 1996; Hähnel et al. 2003; Chui and Rangarajan 2003; Jain et al. 2007). However, because we deal with deformed range data due to movements of the range sensor during the scanning process, conventional rectification and alignment methods for non-rigid models are not appropriate for our objective. We therefore developed new rectification algorithms suitable for the distorted range data taken by the moving range sensor. In this article, we describe two methods to rectify 3D range data.

- 3D alignment (3DA) based method
- Structure from motion (SFM) based method

These methods do not rely on motion sensors: our FLRS combined with software rectification can achieve higher accuracy than hardware solutions. The two methods have slightly different purposes: one emphasizes accuracy and the other emphasizes ease of use. The first method, 3DA-based,

can be applied easily, but the results are less accurate, especially under strong wind. The second method, SFM-based, can achieve high accuracy, but it requires preparation for scanning such as the calibration and synchronization of sensors.

The first method, the 3DA-based method, is an extension of our 3D alignment algorithm, 3DTM (Wheeler and Ikeuchi 1995). This algorithm assumes that the FLRS is usually used in conjunction with a ground-based sensor. It applies to data related to structures that are both visible and invisible from the ground-based sensor, and to data from overlapping regions. The 3DTM algorithm determines rotation and translation parameters and compares data in the overlapping regions of two range images, while the 3DA algorithm determines motion parameters in addition to rotation and translation parameters (Masuda et al. 2005).

The second method, the SFM-based method, uses distorted range data gathered by a moving range sensor and image sequences gathered by a video camera mounted on the FLRS. Combining image sequences and range data, we estimate the sensor motion parameters. This method can obtain camera motion parameters of much greater accuracy than those obtained by a full-perspective factorization (Han and Kanade 1999), with beginning factorization results as initial estimations. Finally, using refined camera motion parameters, the distorted range data are rectified (Banno and Ikeuchi 2005).

We compare the results of our algorithms against a inertial sensor (IS) based method using a regular and an angular accelerometer on the FLRS; verify that the proposed methods achieve high accuracy. Then we compare two proposed methods.

This article is organized as follows: Sect. 2 explains the design issues of our FLRS system. The first method, the 3DA-based method, is presented in Sect. 3. In Sect. 4, we describe the SFM-based rectification method, which utilizes
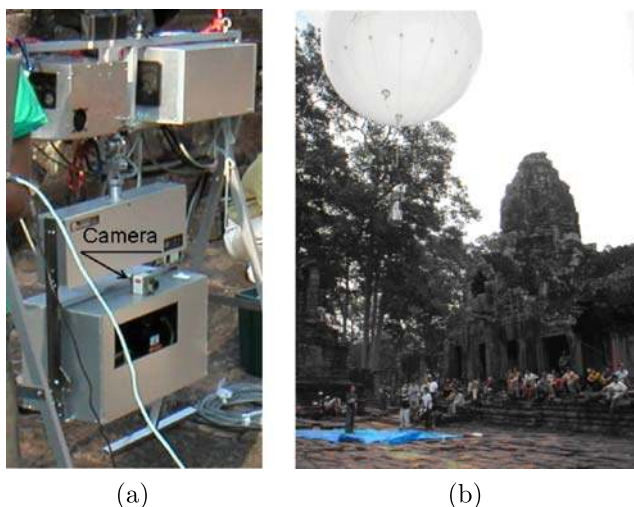
image sequence and range data. Experimental results by both methods are described in Sect. 5, in which we compare the two methods as well as the IS-based method. Finally, in Sect. 6 we present our conclusions as well as some digital photographs of the Bayon temple created from our 3D data.

2 Flying Laser Range Sensor Hardware

The FLRS (Flying Laser Range Sensor) was developed to measure large-scale objects from the air by using a balloon as a base rather than constructing scaffolds. With respect to the measurement principle, the passive stereopsis method could capture images without the influence of balloon motion. However, its results would lack the accuracy needed for the preservation and repair of cultural heritage objects. The laser radar method is suitable for outdoor measurement of large objects, and we therefore adopted a laser range sensor that uses the "time-of-flight" principle. Figure 5 shows a close-up of the FLRS and a scene showing the measurement process being carried out in the Bayon temple. A video camera is mounted on the platform near the range sensor (Fig. 5a) to check the scanning area and to stock the image sequences.

We have two types of FLRSs, which have different range capacities. Each FLRS is composed of a scanner unit, a controller and a personal computer (PC). These three units are suspended beneath a balloon.

The scanner unit includes a laser range finder specially designed to be suspended from a balloon. Figure 6 shows the interior of the scanner unit. It consists of a spot laser radar unit and two mirrors. We chose the LARA25200 and LARA53500 supplied by Zoller+Fröhlich GmbH (http://www.zf-lase.com/) as laser radar units because of their high sampling rate. Each laser radar unit is mounted on an FLRS

scanner unit. The two systems equipped with LARA25200 and LARA53500 are respectively referred to as the "25 m sensor" and the "50 m sensor".

There are two mirrors inside each unit to give direction to the laser beam. One is a polygonal mirror with four reflection surfaces, which determines the azimuth of the beam. In normal use, this mirror, which rotates rapidly, controls the horizontal direction of the laser beam. The other is a plane mirror (swing mirror) that determines the elevation of the laser beam. The plane mirror swings slowly to control the vertical direction of the laser beam. Combining two mirrors, both sensors take a range image in a one-second measurement with more pixels along the horizontal direction (900 pixels) than vertical direction (160 pixels). The view angle of this image is 45° horizontal and 30° vertical. In addition, the 25 m sensor takes a range image of 900×800 pixels with the same view angle in a five-second measurement. The specifications of two units are shown in Table 1.

Figure 6 shows the interior view of the 25 m sensor. The laser beam emitted from the LARA is first directed to a surface of the polygonal mirror. Then the polygonal mirror reflects the laser beam into the plane mirror. The plane mirror also reflects the beam into the outside of the unit (lower part of Fig. 6).

Fig. 5 FLRS. **a** Close-up. **b** Measurement scene
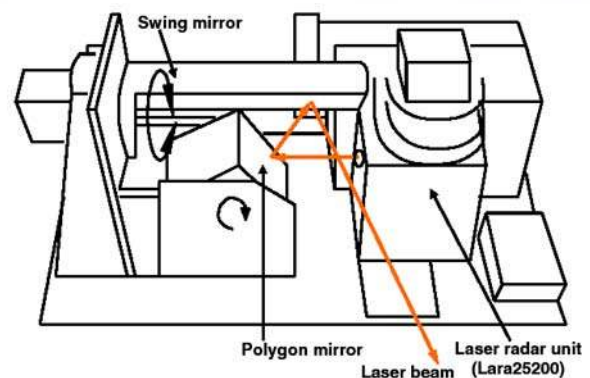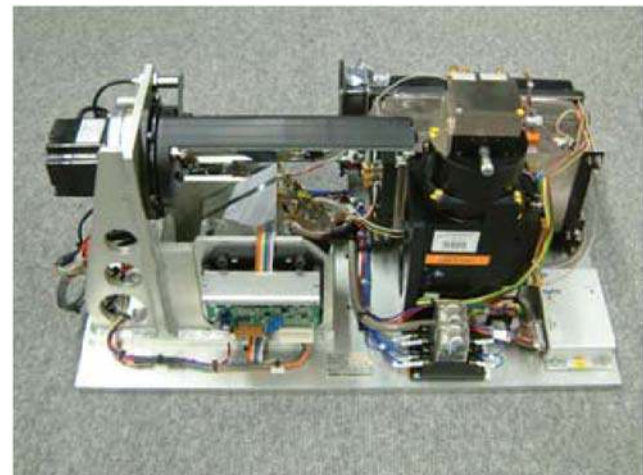
Fig. 6 Interior of a scanner unit (25 m sensor)

Table 1 The specifications of the 25 m and 50 m sensors

	25 m Sensor	50 m Sensor
Ambiguity interval	25.2 m	53.5 m
Minimum range	1.0 m	1.0 m
Resolution	1.0 mm	1.0 mm
Sampling rate	$\leq$625,000 pixels/s	$\leq$500,000 pixels/s
Linearity error	$\leq$ 3 mm	$\leq$5 mm
Range noise at 10 m	$\geq$1.0 mm	$\geq$ 1.5 mm
Range noise at 25 m	$\geq$1.8 mm	$\geq$ 2.7 mm
Laser output power	23 mW	32 mW
Laser wavelength	780 nm	780 nm
Angle Resolution		
Horizontal	0.05°	0.05°
Vertical	0.02°	0.02°
Horizontal field	$\leq$90°	$\leq$90°
Vertical field	$\leq$ 30°	$\leq$30°
Scanning period/range image	$\leq$15 s	$\leq$1 s

2.1 Controller Unit and PC

The controller is composed of a signal processing unit, an interface unit, a mirror controller, and a power supply unit. The signal processing unit receives the signals from the PC and performs actual control of rotation angles of the mirrors and the laser radar unit. The range data obtained by the laser radar and the angle data obtained by the mirror encoders are synchronized and combined in the interface board.

The PC includes a CPU board, an image capture board, and several PCI boards. Main commands for the mirror operations and the laser on/off are sent through the LAN cable between the PC and the controller unit. The synchronized range data and encoder data set are transmitted to the PC via one of the PCI boards. The data set is stored in the PC and converted into 3-dimensional shape data. The PC on the balloon platform is operated remotely via another laptop PC on the ground through a LAN cable.

2.2 Operation

The balloon is filled with helium gas; the diameter of the balloon is 5.0 meter and its maximum buoyancy is about 60 kgf. The balloon is made of a particularly flexible chloroethene, which avoids rapid expansion of a hole in an emergency.

Floating in the air, the balloon is controlled by several hands on the ground with four pieces of rope. To scan an objective region, the FLRS is broadly moved to a nearby site by humans. The facing direction of the FLRS is controlled by the pan and tilt mechanism, which can point the scanner unit from the horizontal direction to the area directly below,

with a scope of 180 degrees from side to side. Monitoring the video images, we can adjust the direction.

The combination of the polygonal mirror's rotation and the plane mirror's swing makes up a range image that alternates between a raster scan order and its reverse order. This mechanism can record the depth value and the timing of each pixel in the range image because the sampling rate of the laser radar is known. In addition, the video camera on the FLRS is synchronized with the laser radar; the image capture board on the PC begins to stock the image sequence in 30 fps simultaneously with the start of the scanning process.

3 3D Alignment with Deformation Parameters

In measuring large-scale objects, we can utilize some range data sets taken by another range sensor fixed on the ground. The ground-base sensors can measure many of the regions; our FLRS was originally devised to complement the measurement for regions that are invisible from the ground. Some parts of a range image taken by the FLRS are also taken by another range sensor fixed on the ground. Based on these overlapping regions, we derive an algorithm that rectifies the distorted range data obtained by the FLRS, and aligns them to the range data taken by the ground-based sensors.

In the first half of this section, we describe a robust method to estimate the translation and rotation parameters for a conventional non-rigid body alignment. The second half deals with an alignment between a ground-based sensor's data and deformed data. The extended method rectifies the distorted FLRS data and aligns them to the other range data of the ground-based sensor.

3.1 Robust Determination of Translation and Rotation Parameters

Several alignment algorithms that estimate translation and rotation parameters for the registration of rigid bodies have been proposed. Besl and McKey designed an alignment algorithm, referred to as the Iterative Closed Point (ICP) algorithm, of a pair of range data by finding corresponding point pairs and determining rotation and translation parameters to reduce the sum of squared distances (SSD) among these pairs (Besl and McKay 1992). However, this algorithm only handles cases when one data set is a subset of another data set. Zhang extended the alignment algorithm to handle general cases by introducing statistical measures (Zhang 1994). Chen and Medioni proposed to use the distance between a point and plane instead of a point to point pair (Chen and Medion 1992). Wheeler and Ikeuchi employed the Lorentzian distance instead of SSD to avoid the

effect of outliers (Wheeler and Ikeuchi 1995). A recent survey of alignment algorithms can be found in Rusinkiewicz and Levoy (2001).

The original algorithm aligns the closest points of two data sets so as to minimize the distance summation of point-to-point distance (Besl and McKay 1992). The minimization of the error function is represented as follows (Wheeler and Ikeuchi 1995):

$$E(\mathbf{p}) = \frac{1}{N} \sum_i^N \rho(z_i(\mathbf{p})), \tag{1}$$

where

$$\mathbf{p} = (\mathbf{T}, \mathbf{q}), \tag{2}$$

$$z_i(\mathbf{p}) = \|\mathbf{R}(\mathbf{q})\mathbf{x}_i + \mathbf{T} - \mathbf{y}_i\|^2, \tag{3}$$

$$\rho(z_i(\mathbf{p})) = \log\left(1 + \frac{1}{2} z_i(\mathbf{p})\right), \tag{4}$$

$\mathbf{T}$:   translation vector (sensor position),

$\mathbf{R}(\mathbf{q})$:   rotation matrix corresponding to quaternion q,

$\mathbf{x}_i$: ith point in the data set of interest,

$\mathbf{y}_i$:   the corresponding point of x_i in the measured data,

N: the number of data points.

The range images are aligned iteratively by moving (translating/rotating) the measured data according to the estimated parameters. The movement is determined such that the total distance sum between the corresponding points is minimized. As for the rotation matrix, we use the quaternion representation of 3 degrees of freedom. In the direct square summation of error function (3), considerable noise leads to imprecise alignment. In this algorithm, M-estimation (Lorentz function) is used for noise elimination by considering the probability distribution of the error (Wheeler and Ikeuchi 1995). Using this error metric $E(\mathbf{p})$, we compute the parameters $\mathbf{p}$ that fulfill the following equation:

$$\mathbf{p}_{opt} = \arg\min_{\mathbf{p}} E(\mathbf{p}). \tag{5}$$

For the gradient-based solution of non-linear optimization (Press et al. 1988), the descent gradient is computed as follows:

$$\frac{\partial E}{\partial \mathbf{p}} = \frac{1}{N} \sum_i^N \frac{\partial \rho(z_i)}{\partial z_i} \cdot \frac{\partial z_i}{\partial \mathbf{p}}$$

$$= \frac{1}{N} \sum_i^N w(z_i) z_i \frac{\partial z_i}{\partial \mathbf{p}}, \tag{6}$$

where

$$w(z_i) = \frac{1}{z_i} \cdot \frac{\partial \rho(z_i)}{\partial z_i}.$$

If we evaluate $\partial z_i/\partial \mathbf{p}$ by identifying quaternion q_I, we can represent $\partial z_i/\partial \mathbf{p}$ as

$$\frac{\partial z_i(\mathbf{p})}{\partial \mathbf{p}} = 2(\mathbf{R}(\mathbf{q})\mathbf{x}_i + \mathbf{T} - \mathbf{y}_i) \frac{\partial(\mathbf{R}(\mathbf{q})\mathbf{x}_i + \mathbf{T} - \mathbf{y}_i)}{\partial \mathbf{p}}\bigg|_{\mathbf{q}_I}$$

$$= \begin{bmatrix} 2(\mathbf{x}_i + \mathbf{T} - \mathbf{y}_i) \\ -4\mathbf{x}_i \times (\mathbf{T} - \mathbf{y}_i) \end{bmatrix}. \tag{7}$$

The detailed explanation of the derivation with respect to the quaternion is shown in Wheeler (1996).

Our goal is to simultaneously determine deformation, translation, and rotation parameters by comparing the target data to its corresponding data. The translation and rotation parameters are determined in a minimization paradigm. If we fix these parameters, the determination of the deformation parameter becomes an iterative shape-matching problem.

3.2 Simultaneous Determination of Deformation Parameters for FLRS

We extend the parameter estimation of the alignment formulation for the distorted range data obtained by the moving range sensor. While translation and rotation registration is due to the sensor movement among multiple views, shape deformation registration is due to the sensor movement during one scan. We refer to them as inter-scanning registration and intra-scanning registration respectively. In addition to the registration of inter-scanning transformation parameters, we must also determine the intra-scanning movement parameters.

In the case of a short-term measurement, we can suppose that pose and motion change of the FLRS is generated by

- Initial velocity
- Initial angular velocity
- Acceleration generated by external force
- Angular acceleration generated by external moment

We can ignore the influence of translation and angular acceleration because the FLRS needs only one second to scan, and insignificant rotation can be approximated to translation movement. Hence we consider only constant velocity movement. We set up the deformation equation under this assumption.

In this case, the displacement of $\mathbf{x}_i$ by deformation is represented only by using constant velocity vector $\mathbf{v}$ of FLRS movement. To estimate $\mathbf{v}$ as a deformation parameter in the same way as the determination of rigid-body transformation (translation and rotation) parameters, we have only to replace $\mathbf{x}_i$ in (3) with $(\mathbf{x}_i - \tau_i \mathbf{v})$. Here, τ_i is the time passed since the start of the scanning to the capture time of point i.

Let $(\mathbf{x}_i - \tau_i \mathbf{v})$ be $\mathbf{g}(\mathbf{x}_i, \mathbf{v})$, and the descent gradient in (6) can be written, evaluating $\partial z_i / \partial \mathbf{p}$ by identifying quaternion q_I, as follows:

$$
\frac{\partial z_i(\mathbf{p})}{\partial \mathbf{p}} = 2(\mathbf{R}(\mathbf{q})\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{y}_i)
$$

$$
\times \left. \frac{\partial (\mathbf{R}(\mathbf{q})\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{y}_i)}{\partial \mathbf{p}} \right|_{\mathbf{q}_I}
$$

$$
= \begin{bmatrix} 2(\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{y}_i) \\ -4\mathbf{g}(\mathbf{x}_i, \mathbf{v}) \times (\mathbf{T} - \mathbf{y}_i) \\ 2(\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{y}_i) \frac{\partial(\mathbf{g}(\mathbf{x}_i, \mathbf{v}))}{\partial \mathbf{v}} \end{bmatrix} \quad (8)
$$

where

$$
\mathbf{p} = (\mathbf{T}, \mathbf{q}, \mathbf{v}),
$$

$$
\frac{\partial \mathbf{g}(\mathbf{x}_i, \mathbf{v})}{\partial \mathbf{v}} = - \begin{bmatrix} \tau_i & 0 & 0 \\ 0 & \tau_i & 0 \\ 0 & 0 & \tau_i \end{bmatrix}. \quad (9)
$$

This straightforward extension causes unstable convergence of the deformation registration because the obtained translation, rotation, and deformation parameters overreach their optimum if every parameter is applied simultaneously to the deformation: every parameter interferes with the others.

In order to prevent this interference, we redesign our extended formulation to remove the translation and rotation effect caused only by deformation. The basic idea is to recover the position and posture that change due to deformation. This is implemented by the preliminary rigid-body transformation that determines only the deformation parameter. First, all parameters $\mathbf{T}$, $\mathbf{q}$ and $\mathbf{v}$ are acquired by (6) and (8). Then the preliminary rigid-body transformation is determined only by the deformation parameter $\mathbf{R}_o$ and $\mathbf{T}_o$ as follows:

$$
\mathbf{g}'(\mathbf{x}_i, \mathbf{v}) = \mathbf{R}_o \mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T}_o, \quad (10)
$$

where

$$
(\mathbf{R}_o, \mathbf{T}_o) = (\mathbf{R}(\mathbf{q}_o), \mathbf{T}_o),
$$

such that

$$
(\mathbf{q}_o, \mathbf{T}_o) = \arg \min_{\mathbf{q}, \mathbf{T}} \sum_i^N \| \mathbf{R}(\mathbf{q}) \mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{x}_i \|^2. \quad (11)
$$

$\mathbf{q}_o$, $\mathbf{T}_o$ can be derived from the following equation:

$$
\frac{\partial \sum_i^N \epsilon_i^2}{\partial \mathbf{P}_o} = \sum_i^N 2\epsilon_i \cdot \frac{\partial \epsilon_i}{\partial \mathbf{P}_o} = \mathbf{0}, \quad (12)
$$

where

$$
\epsilon_i = \mathbf{g}'(\mathbf{x}_i, \mathbf{v}) - \mathbf{x}_i = \mathbf{R}(\mathbf{q}_o)\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T}_o - \mathbf{x}_i,
$$

$$
\mathbf{P}_o = (\mathbf{q}_o, \mathbf{T}_o).
$$

This is a conventional registration problem, but it is unnecessary to strictly solve the above equation. In fact, $\mathbf{R}_o$ does not affect the stable convergence so much as $\mathbf{T}_o$. If $\mathbf{R}_o$ is ignored, $\mathbf{T}_o$ in (12) is concretely derived as follows:

$$
\sum_i^N (\mathbf{g}(\mathbf{x}_i, \mathbf{v}) + \mathbf{T}_o - \mathbf{x}_i) = \mathbf{0}
$$

$$
\therefore \mathbf{T}_o = -\frac{\sum_i^N (\mathbf{g}(\mathbf{x}_i, \mathbf{v}) - \mathbf{x}_i)}{N} = \frac{\sum_i^N \tau_i}{N} \mathbf{v}. \quad (13)
$$

Equation (3) is replaced with:

$$
z_i(\mathbf{p}) = \| \mathbf{R}(\mathbf{q}) \mathbf{g}'(\mathbf{x}_i, \mathbf{v}) + \mathbf{T} - \mathbf{y}_i \|^2. \quad (14)
$$

Equations (10) and (9) are modified into:

$$
\mathbf{g}'(\mathbf{x}_i, \mathbf{v}) = \mathbf{x}_i - \left(\tau_i - \frac{\sum_i^N \tau_i}{N} \right) \mathbf{v}, \quad (15)
$$

$$
\frac{\partial \mathbf{g}'(\mathbf{x}_i, \mathbf{v})}{\partial \mathbf{v}}
$$

$$
= - \begin{bmatrix} \tau_i - \frac{\sum_i^N \tau_i}{N} & 0 & 0 \\ 0 & \tau_i - \frac{\sum_i^N \tau_i}{N} & 0 \\ 0 & 0 & \tau_i - \frac{\sum_i^N \tau_i}{N} \end{bmatrix}. \quad (16)
$$

In this method, the acquisition of a good initial parameter is significant for the optimal registration result. In our implementation, the initial transformation parameter is set manually by GUI with enough accuracy to reach the optimum.

4 Rectification by Combining Image Sequence and Range Data

The second method, the SFM-based method, utilizes range data and an image sequence obtained by a video camera mounted on the FLRS as follows: First, the motion of the FLRS is roughly estimated only by the obtained images; we use a full-perspective factorization (Christy and Horaud 1996; Han and Kanade 1999). These approximate motion parameters are utilized as an initial solution. Then the motion parameters for rectification are estimated based on an optimization imposing some constraints, which include information derived from the image sequence and the distorted range data. For the sake of an accurate estimation of motion parameters, the optimization consists of three constraints:

range data, balloon motion smoothness, and tracking of interest points. The camera motion parameters can be found through the minimization of a global functional. Finally, by using the camera motion parameters, the distorted range data are rectified.

4.1 Problem Setting

We assume that F images are given by the video camera while obtaining a single range image; the period is defined as T_s. Then P interest points are tracked over all the frames; each 3D interest point p is expressed as $\mathbf{S}_p$ in the world coordinate system; the world coordinate system is defined by the full-perspective factorization.

The 3D point p is projected on the image plane at $(u_p(t), v_p(t))^t$ at time t. We can obtain these projected points at the discrete time t_f $(1 \le f \le F)$ by using an interest point detector. The relation between the discrete time t_f and the frame number f is described as $t = f/30$ because we use a video camera of 30 fps. The positions of these projected points as a continuous time are determined using an interpolation technique.

By using the intrinsic camera parameter matrix $\mathbf{A}$, the image projection is described as follows:

$$\begin{pmatrix} u_p(t) \\ v_p(t) \\ 1 \end{pmatrix} \propto \mathbf{A} \begin{pmatrix} \mathbf{i}(t)^t \\ \mathbf{j}(t)^t \\ \mathbf{k}(t)^t \end{pmatrix} (\mathbf{S}_p - \mathbf{T}(t)), \tag{17}$$

t:	scanning time $0 \le t \le T_s$,
f:	frame number $1 \le f \le F$,
p:	the number of the interest points $1 \le p \le P$,
$\mathbf{T}(t)$:	camera position at t,
$\mathbf{S}_p$:	the 3D position of p-th interest point,
$\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$:	the x, y and z axis of the camera coordinate system at t.

The range sensor outputs a cloud of 3D points; each point is converted into the camera coordinate description $\mathbf{x}(t) = (x(t), y(t), z(t))^t$ because the range sensor and the camera are calibrated.

4.2 Motion Estimation by Combining Range Data and Image Sequence

We adopt an optimization strategy to estimate motion parameters for the correct rectification. With respect to set the cost function to minimize, we impose three constraints that combine the image sequence and the range data. The first is a range data constraint, which utilizes temporally geometric relations between 3D points and the camera. The second is a smoothness constraint, which guarantees a smooth motion of the balloon. The third is a tracking constraint, which is known as the bundle adjustment.

4.2.1 Range Data Constraint

The range sensor outputs a 3D point $\mathbf{x}(t)$ at t. Since the range sensor and the video camera are calibrated, we can determine the position of the projected point of $\mathbf{x}(t)$ on the image plane. Let us define the 2D projected point as $\mathbf{u}(t)$. If the point $\mathbf{u}(t)$ matches any interest point $(u_p(t), v_p(t))^t$ on the image at time t_p, we can judge that the range sensor scans the 3D interest point p at time t_p. There are few points that are scanned by the range sensor but invisible from the video camera because the video camera is located near the laser radar (about 10 cm) and these points are more than 5 meters away from the sensor in actual case.

Therefore, we can derive the first constraint as follows:

$$F_A = \sum_{p=1}^{P} \alpha_p \|\mathbf{x}(t_p) - \mathbf{R}^t(t_p)(\mathbf{S}_p - \mathbf{T}(t_p))\|^2 \tag{18}$$

where

$$\alpha_p = \begin{cases} 1 & \text{if point } p \text{ is scanned by the range sensor,} \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbf{R}(t) = (\mathbf{i}(t)\mathbf{j}(t)\mathbf{k}(t))$ means the rotation matrix of the camera pose at t. As described above, t_p is the scan time when the 3D interest point p is scanned by the range sensor. Each t_p is estimated as follows:

$$t_p = \arg\min_t \left\| \mathbf{u}(t) - \begin{pmatrix} u_p(t) \\ v_p(t) \end{pmatrix} \right\|. \tag{19}$$

It is not always possible to use all interest points in images for this constraint. Several interest points in images are on 3D corners and 3D edges, which cause mismatches in the range image. In order to exclude these interest points on 3D corners, we set $\alpha_p = 0$ in (18) for the points whose neighboring points in the range image have a variance larger than a threshold. The interest points, which are located out of the laser's range, are also excluded.

4.2.2 Smoothness Constraint

One of the most significant reasons for adopting a balloon platform is to be free from the high frequency that occurs with a helicopter platform. A balloon platform is only under the influence of low frequency: the balloon of our FLRS is held with some wires swayed only by wind. This means that the movement of the balloon is expected to be smooth. Certainly, the movement of the balloon is free from rapid acceleration, rapid deceleration, or acute change of course. Taking this fact into account, we consider the following function:

$$F_B = \int_0^{T_s} \left(w_1 \left\| \frac{\partial^2 \mathbf{T}(t)}{\partial t^2} \right\|^2 + w_2 \left\| \frac{\partial^2 \mathbf{q}(t)}{\partial t^2} \right\|^2 \right) dt. \tag{20}$$

Here, $\mathbf{T}(t)$ denotes the position of the camera; w_1, w_2 are weighted coefficients; and $\mathbf{q}(t)$ is a unit quaternion that represents the camera pose. The first term of the above integrand represents smoothness with respect to the camera's translation while the second one represents smoothness with respect to the camera's rotation. When the motion of the camera is smooth, the function F_B has a small value.

4.2.3 Tracking Constraint

Any interest point $\mathbf{S}_p$ must be projected near the observed interest points $(u_p(t_f), v_p(t_f))$ on each frame f. This constraint is well known as Bundle Adjustment (Brown 1976). When the structure, motion, and shape have been roughly obtained, this technique is utilized to refine them through an image sequence.

From (17), the following relationship is conducted:

$$\mathbf{A}^{-1} \begin{pmatrix} u_p(t_f) \\ v_p(t_f) \\ 1 \end{pmatrix} \propto \begin{pmatrix} \mathbf{i}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_t)) \\ \mathbf{j}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_t)) \\ \mathbf{k}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_t)) \end{pmatrix}. \tag{21}$$

Supposing $(\hat{u_p}(t_f), \hat{v_p}(t_f), 1)^t = \mathbf{A}^{-1}(u_p(t_f), \; v_p(t_f), \; 1)^t$, we set the third constraint:

$$F_C = \sum_{f=1}^{F} \sum_{p=1}^{P} \left(\left(\hat{u_p}(t_f) - \frac{\mathbf{i}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_f))}{\mathbf{k}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_f))} \right)^2 \right.$$
$$\left. + \left(\hat{v_p}(t_f) - \frac{\mathbf{j}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_f))}{\mathbf{k}(t_f)^t \cdot (\mathbf{S}_p - \mathbf{T}(t_f))} \right)^2 \right). \tag{22}$$

The minimization of F_C leads to the correct tracking of the interest points by a moving camera.

4.2.4 The Global Cost Function

Based on the above considerations, it will be found that the next cost function should be minimized. Consequently, the weighted sum

$$F = w_A F_A + w_B F_B + w_C F_C \tag{23}$$

leads to the global cost function. The coefficients w_A, w_B and w_C are determined experimentally so that three terms take almost the same magnitude. Measuring $\mathbf{x}(t)$, $\mathbf{S}_p$ and $\mathbf{T}(t)$ in the metric system and $(u_p(t), v_p(t))$ in pixel, we set three coefficients as $(w_A, w_B, w_C) = (10^2, 10^5, 10^0)$ and two coefficients in (20) as $(w_1, w_2) = (10^0, 10^1)$ in this study.

To minimize this function, we employ a conjugate gradient method (Polak 1971; Jacobs 1977; Stoer and Bulirsh 1980; Press et al. 1988) to find the search directions. Then, we use the golden section search to determine the minimum along the directions. For optimization, the Levenberg-Marquardt method (Marquardt 1963) is generally employed

to minimize a functional value. The Levenberg-Marquardt method is very effective in estimating a function's parameters, especially to fit a certain function. However, in our function, it is not a parameter fitting problem to minimize the value of F_B, but a simple decreasing of F_B. Therefore we adopt the conjugate gradient method.

4.3 Initial Estimation

To minimize the value of (23), we need a proper initial guess. With an improper initial value, the minimization will be trapped into local minimums. In this study, we utilize the solution by full-perspective factorization (Christy and Horaud 1996; Han and Kanade 1999) as the initial value to minimize the global cost function. Using the weak-perspective projection model, the full perspective factorization iteratively estimates the shape and the camera motion under the perspective model.

For the factorization, we need P interest points tracked from F frames. There are several methods to derive interest points of images (Moravec 1977; Smith and Brady 1997; Lowe 2004). Among them, we adopt Harris operator (Harris and Stephens 1988) for derivation of interest points. Over 300 interest points are derived from each frame; identified points are connected by using window matching. In order to reject outliers, we impose the Epipolar constraint for all adjacent frame pairs. Moreover, we adopt RANSAC (Fischler and Bolles 1981) to estimate all F-Matrices.

5 Experimental Results

We show the experimental results of two data sets; these data sets were obtained in the case of a balloon with a moderate motion and a wide motion. We compared the rectified shapes resulting from our two proposed methods and a IS-based method. Our FLRS system is equipped with a regular and an angular accelerometer. To demonstrate the advantage of our methods, we utilized the rectified shape produced by these physical devices.

In all the experiments, it took one second to scan a range image of 900×160 pixels; the horizontal view angle is $45°$ and the vertical angle is $30°$. The video camera stocks 70 images with 30 fps including the one second of the scanning period. For calibration of the range sensor and the video camera, we utilized the range data obtained by the FLRS fixed on the ground.

Case 1

The data set of Case 1 was obtained when the balloon motion was moderate. Figure 7a shows a photograph of the

Fig. 7 Case 1: **a** The object for the experiment. **b** The 3D model by the ground-base sensor

(a) (b)
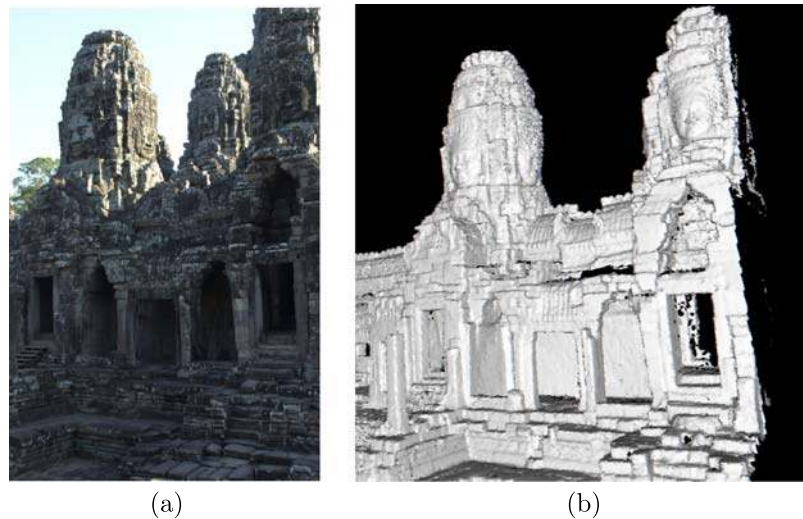
Fig. 8 Case 1: **a** The original distorted range data by FLRS. **b** The range data rectified by the IS-based method. **c** The range data rectified by the 3DA-based method. **d** The range data rectified by the SFM-based method


(a) (b)

(c) (d)
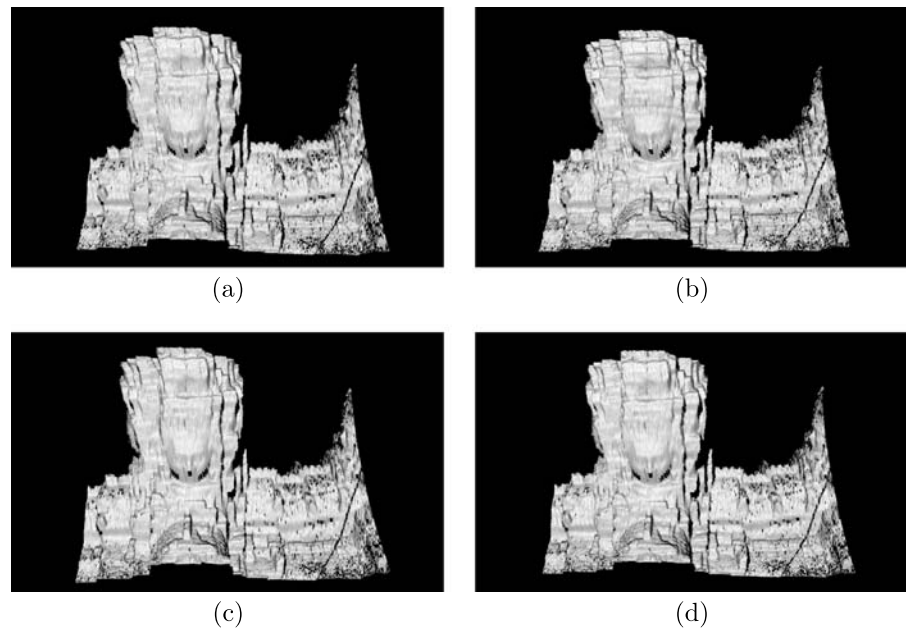
scanned area. The dense model of Fig. 7b is obtained by the Cyrax2500 (http://www.leica-geosystems.com/) fixed on the ground. We treat this model as the correct data hereafter.

The original distorted range data and some rectified shapes are shown in Fig. 8. Figure 8a shows the original distorted range data obtained by the FLRS. The shape of Fig. 8b is the rectified shape based on the measurements of the acceleration meter and the angular accelerometer. Figure 8c is the shape rectified by the 3DA-based method. To produce the rectified shape of Fig. 8c, we utilized the shape of Fig. 7b. Figure 8d is the shape rectified by the SFM-based method of combining the range data and the image sequences of the FLRS.

These shapes in Fig. 8 do not clearly demonstrate the effectiveness of the rectifications due to a moderate balloon motion. To determine these, we evaluated the validity of the rectified shapes numerically, using geometric similarities between the correct shape and each rectified shape.

The results are shown in Table 2. After aligning each rectified model to the correct one, the distances between the corresponding pairs are calculated. The values in Table 2 show the percentages of the pairs with closer distances than the thresholds among all pairs. We set up three thresholds as 1.0 cm, 5.0 cm and 10.0 cm. When we scanned objects using the FLRS fixed on the ground, the accuracy was estimated at about 1.0 cm. We think that a 3D model with accuracy of

Fig. 9 Case 1: Comparisons with the range data obtained by the ground-based sensor. The white regions indicate where the distances of the corresponding pairs are less than 1.0 cm. **a** The original distorted range data by FLRS. **b** The range data rectified by the IS-based method. **c** The range rectified by the 3DA-based method. **d** The range data rectified by the SFM-based method

(a)

(b)

(c)

(d)
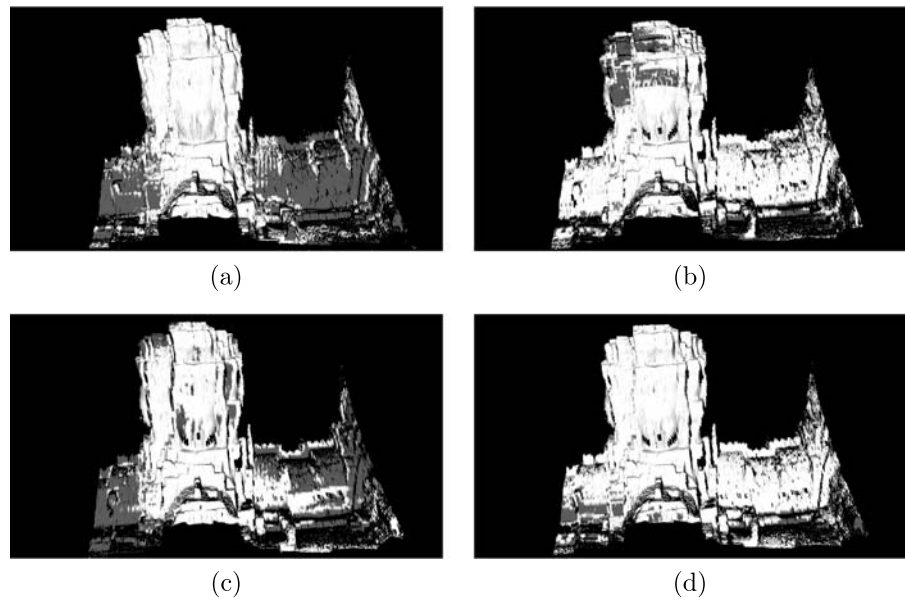
Table 2 The evaluation of the rectified models (Case 1). The percentages of the pairs with closer than the thresholds

	Original	IS-based	3DA-based	SFM-based
≤1.0 cm (%)	58.08	67.23	60.20	80.42
≤5.0 cm (%)	76.45	93.16	89.93	96.18
≤10.0 cm (%)	79.28	97.74	97.13	98.49

less than 5.0 cm would be sufficient for modeling cultural assets.

More than half of the points in the original distorted range data have a 1.0 cm accuracy since this data set was obtained under moderate balloon motion. Using three rectification methods, the areas with 1.0 cm accuracy are increased. The IS-based method can rectify the distorted range data; the advantage of this method is its rapidity because it does not need any iterative calculations. In this data set, the IS-based method shows high performance. The 3DA-based method does not increase the points with 1.0 cm accuracy; nevertheless, this method increases the points with 5.0 cm accuracy almost ninety percent of the time. The SFM-based method demonstrates even greater effectiveness than the IS-based method: more than eighty percent of points have 1.0 cm accuracy.

The appearances of the regions with 1.0 cm accuracy are shown in Fig. 9. The white regions indicate the areas of high accuracy of 1.0 cm.

Figure 10 shows two cross-sections of four models: the correct model by Cyrax2500, the one rectified by the IS-based method, the one rectified by the 3DA-based method, and the one rectified by the SFM-method.

Table 3 The evaluation of the rectified models (Case 2)

|	Original	IS-based	3DA-based	SFM-based
≤1.0 cm (%)	26.30	22.55	35.33	54.30
≤5.0 cm (%)	55.97	45.47	65.18	86.54
≤10.0 cm (%)	70.19	58.13	77.10	93.58

Case 2

The data set of Case 2 was, on the other hand, obtained when the balloon moved widely. Figure 11a shows a photo picture of the scanned area. The dense model of Fig. 7b is the correct data by the Cyrax2500.

Figure 12a is the original distorted shape obtained by the FLRS. It is found that the shape is widely deformed. In Fig. 12b, the shape rectified by the IS-based method is shown. Unlike the previous case, this rectification method leads to a noticeable deformation. There is almost no variation in appearance of Fig. 12c. On the other hand, a glance at Fig. 12d shows that the SFM-based method seems to rectify the shape properly.

Table 3 shows the numerical evaluations in this data set. The wide balloon motion caused a quarter of the area of 1.0 cm accuracy in the original range data. This value clarifies numerically that the IS-based method makes rectification less efficient. In the case of a wide and rapid balloon motion, it is found that the IS-based method is improper to rectify the distorted FLRS data. The 3DA-based method increases the area of 1.0 cm accuracy, but it could not reach the level attained by the SFM-based method. Using the SFM-based method, more than half the area has a 1.0 cm accuracy,
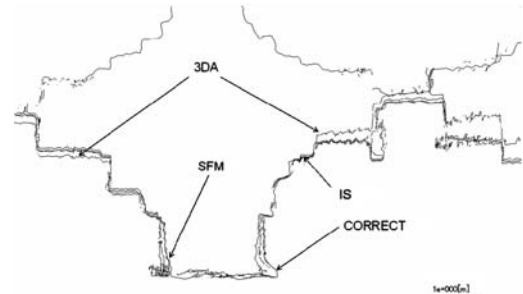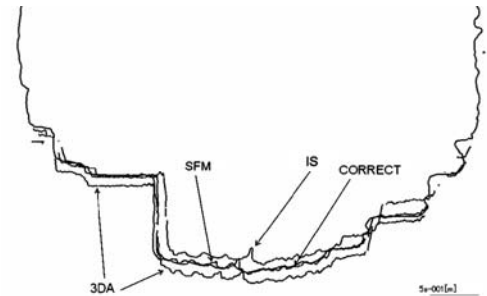
Fig. 10 Case 1: Cross sections of four models; "CORRECT": the correct model by Cyrax2500. "IS": the range data rectified by the IS-based method. "3DA": the range data rectified by the 3DA-based method. "SFM": the range data rectified by SFM-based method
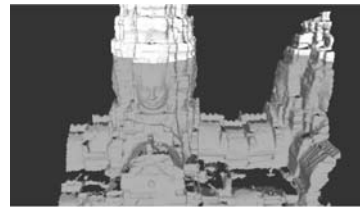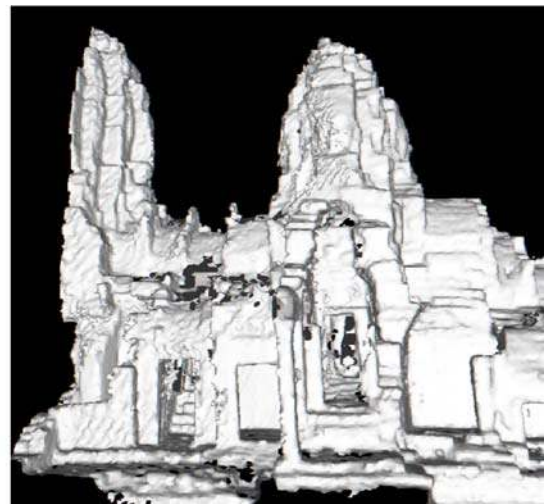

Cutting plane 1
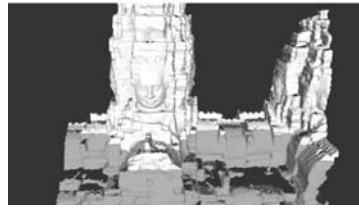
Cutting plane 2

Cross sections by the plane 2

Fig. 11 Case 2: **a** The object for the experiment. **b** The 3D model by the ground-base sensor

(a)   (b)

while more than eighty percent of points have a 5.0 cm accuracy. We think the reason why the 3DA-based method could not achieve the performance by the SFM-based method is as follows: the assumption of the constant velocity movement of the balloon does not hold true in this case. On the other hand, the 3DA-based method works out in the case of moderate balloon motion.

The appearances of the regions of the 1.0 cm accuracy in Case 2 are shown in Fig. 13.

Figure 14 shows two cross sections of four models in Case 2.

Additional Case (Case 3)

We demonstrate an additional data set, in which it takes five seconds to scan a range image of 900×800 pixels; similarly the horizontal view angle is $45°$ and the vertical angle is $30°$. Since it was impossible to rectify the distorted data by the IS-based and 3DA-based methods, we show the range data rectified only by the SFM-based method.

Figure 15 shows the scene of Case 3 and the correct range data by the ground-base sensor. Although the original data

Fig. 12 Case 2: **a** The original distorted range data by FLRS. **b** The range data rectified by the IS-based method. **c** The range data rectified by the 3DA-based method. **d** The range data rectified by the SFM-based method

(a)

(b)

(c)

(d)
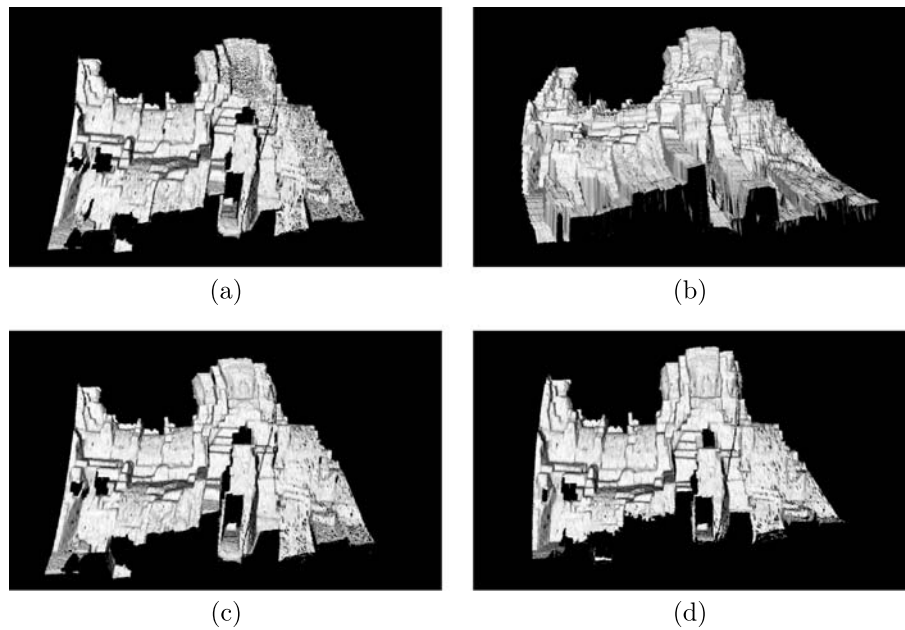
Fig. 13 Case 2: Comparisons with the range data obtained by the ground-based sensor. The white regions indicate where the distances of the corresponding pairs are less than 1.0 cm. **a** The original distorted range data by FLRS. **b** The range data rectified by the IS-based method. **c** The range data rectified by the 3DA-based method. **d** The range data rectified by the SFM-based method

(a)

(b)

(c)

(d)
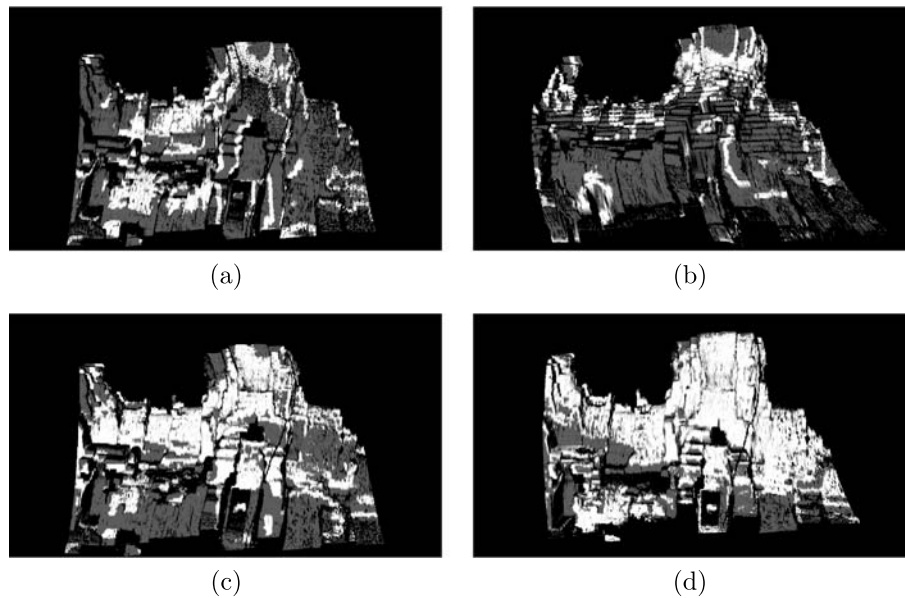
are widely deformed due to the five-seconds scanning, it is found that the SFM-based method rectifies the data properly.

The original distorted range data and the shape rectified by the SFM-based method are shown in Fig. 16.

The appearances of the regions of 1.0 cm accuracy in Case 3 are shown in Fig. 17, which indicates the high performance of the SFM-based method.

6 Conclusions

We have described our novel sensor, FLRS, for measuring large-scale objects. The FLRS is a new-sprung system in our ongoing project to digitize the Bayon temple at the Angkor ruin, in Cambodia. The FLRS can digitize objects from the air while suspended from the underside of the balloon platform. The advantage of the FLRS over conventional sensors is that the FLRS can easily measure the regions invisible from the ground. We have also described two kinds of rectification methods to solve the unique problem for the FLRS: the distorted range data due to the balloonfs motion. One is the 3DA-based method, which rectifies the distorted range data by using another range data set obtained from the ground-based sensor; another is the SFM-based method, which estimates the sensor motion combining the range data and the image sequences. We have confirmed the effective-

Fig. 14 Case 2: Cross sections of four models; "CORRECT": the correct model by Cyrax2500. "IS": the range data rectified by the IS-based method. "3DA": the range data rectified by the 3DA-based method. "SFM": the range data rectified by the SFM-base method


Cutting plane 1

Cross sections by the plane 1

Cutting plane 2

Cross sections by the plane 2
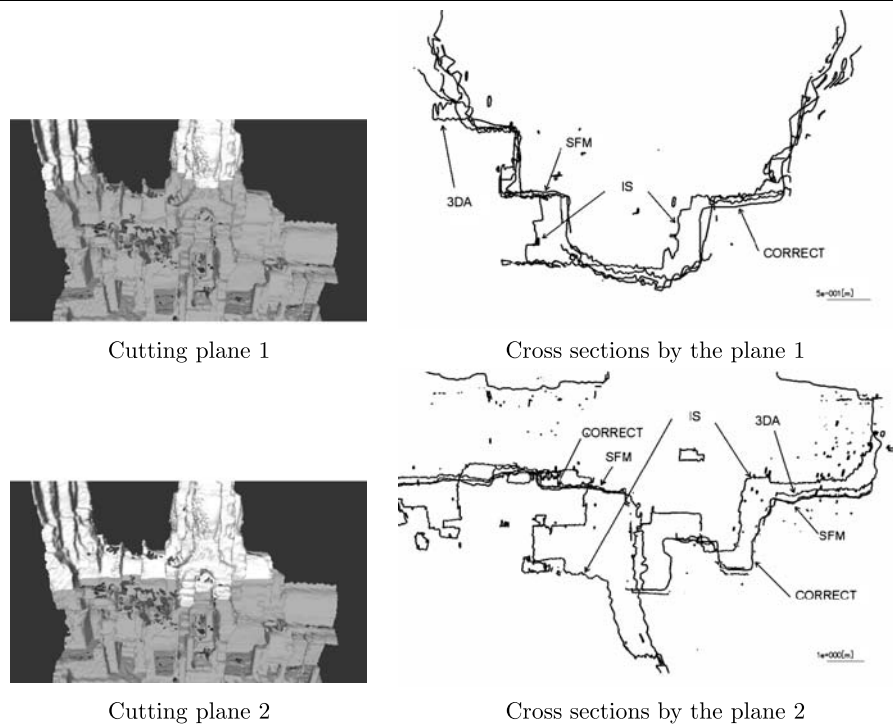
Fig. 15 Case 3: **a** The object for the experiment. **b** The 3D model by the ground-base sensor


(a)

(b)
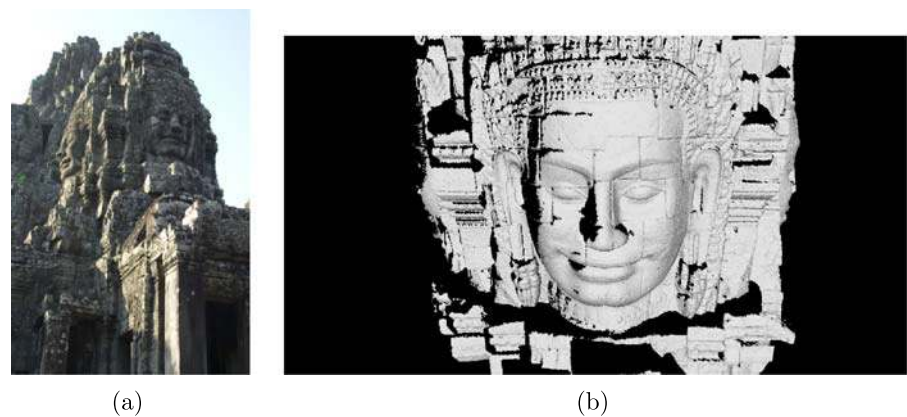
Fig. 16 Case 3: **a** The original distorted range data by FLRS. **b** The range data rectified by the SFM-based method

(a)

(b)
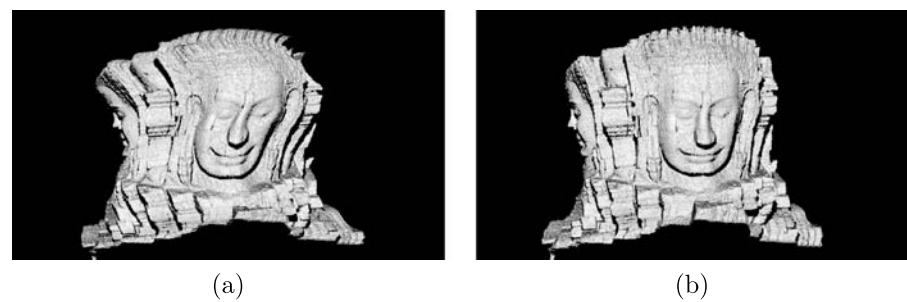
Fig. 17 Case 3: Comparisons with the range data obtained by the ground-based sensor. The white regions indicate where the distances of the corresponding pairs are less than 1.0 cm. **a** The original distorted range data by FLRS. **b** The range data rectified by the SFM-based method

(a)

(b)
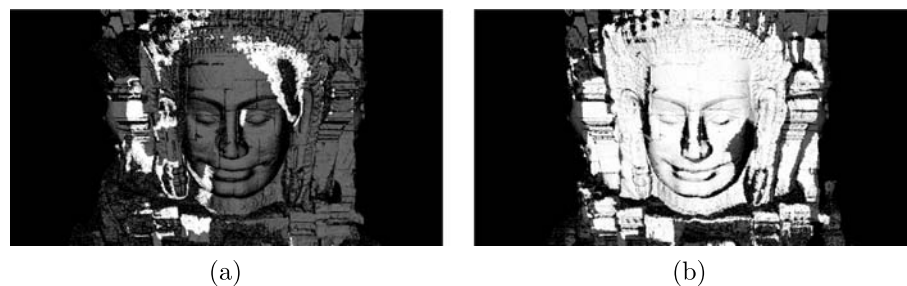
Fig. 18 The overview of the
"Digital Bayon"
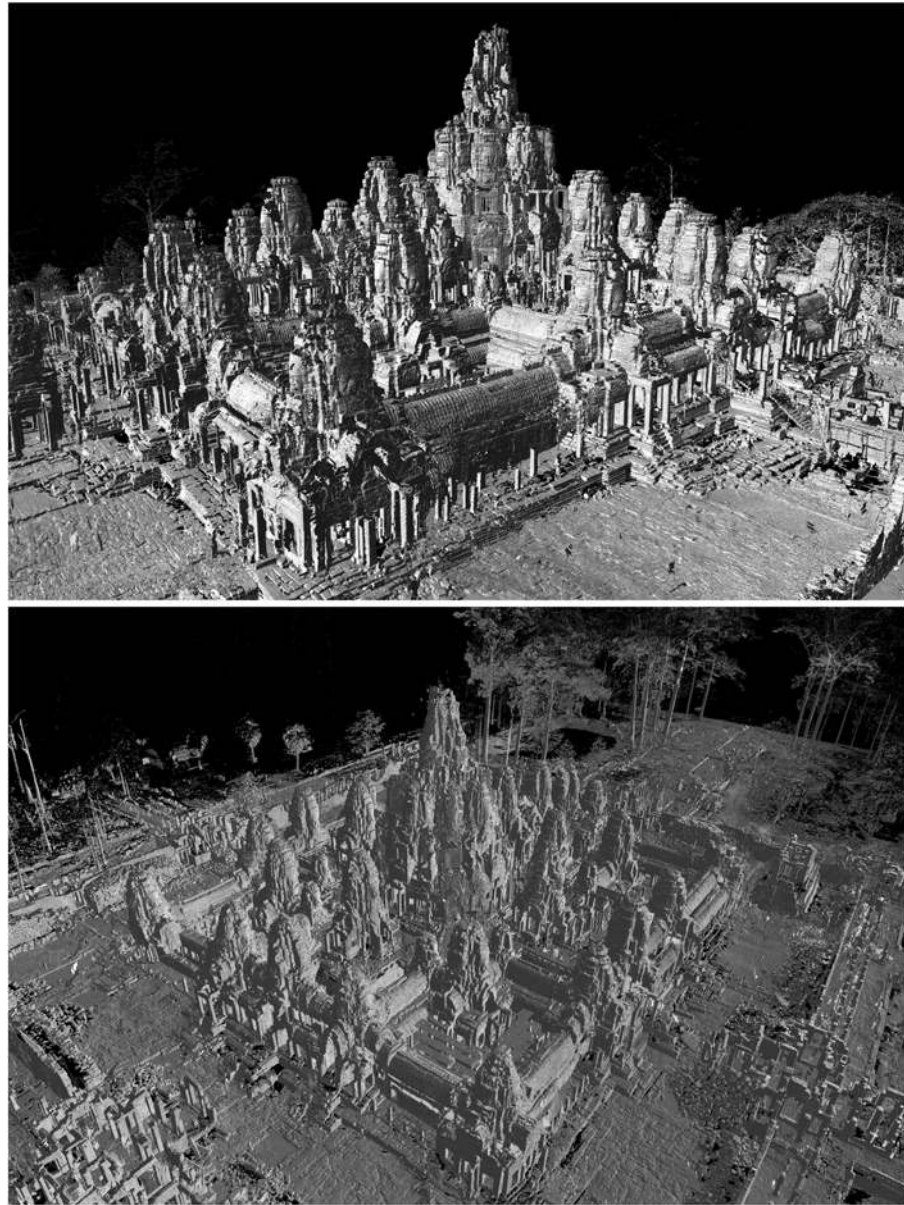
ness of these methods by comparing the IS-based rectifica-
tion method. Among them, the SFM-based method achieved
higher accuracy for the rectified range data.

Finally, Fig. 18 shows the measurement result recon-
structed using all range data obtained by all sensors, FLRSs,
Cyrax, Z + F Imager and Climbing Sensor (Matsui et al.
2005). This 3D model consists of 20,000 range images ob-
tained in all missions and the total size is about 200 GByte.[1]

Measurement missions in the Bayon temple were conducted in coop-
eration with the Japanese government team for Safeguarding Angkor
(JSA).

References

Banno, A., & Ikeuchi, K. (2005). Shape recovery of 3D data ob-
tained from a moving range sensor by using image sequences. In
*Proceedings of the international conference on computer vision
(ICCV2005)* (Vol. 1, pp. 792–799).

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-
D shapes. *IEEE Transactions on Pattern Analysis and Machine
Intelligence*, *14*, 239–256.

Brown, D. (1976). The bundle adjustment—progress and prospect. In
XIII congress of the ISPRS, Helsinki.

[1]http://www.cvl.iis.u-tokyo.ac.jp/research/bayon/.

Chen, Y., & Medion, G. (1992). Object modeling by registration of multiple range images. *Image and Vision Computing*, *10*(3), 145–155.

Christy, S., & Horaud, R. (1996). Euclidean shape and motion from multiple perspective views by affine iterations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(11), 1098–1104.

Chui, H., & Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, *89*, 114–141.

Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of SIGGRAPH'96* (pp. 303–312). New York: ACM.

Feldmar, J., & Ayache, N. (1996). Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, *18*(2), 99–119.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

Hähnel, D., Thrun, S., & Burgard, W. (2003). An extension of the ICP algorithm for modeling nonrigid objects with mobile robots. In *Proceedings of the international joint conference on artificial intelligence (IJCAI)*.

Han, M., & Kanade, T. (1999). *Perspective factorization methods for euclidean reconstruction* (Technical report: CMU-RI-TR-99-22). Robotics Institute, Carnegie Mellon University.

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of Alvey vision conference* (pp. 147–152).

Hirota, Y., Masuda, T., Kurazume, R., Ogawara, K., Hasegawa, K., & Ikeuchi, K. (2004). Designing a laser range finder which is suspended beneath a balloon. In *Proceedings of the 6th Asian conference on computer vision (ACCV2004)* (Vol. 2, pp. 658–663).

Ikeuchi, K., Nakazawa, A., Hasegawa, K., & Ohishi, T. (2003). The Great Buddha project: Modeling cultural heritage for VR systems through observation. In *Proceedings of the 2nd IEEE and ACM international symposium on mixed and augmented reality (ISMAR2003)*.

Jacobs, D. A. (1977). *The state of the art in numerical analysis*. London: Academic Press.

Jain, V., Zhang, H., & van Kaick, O. (2007, accepted). Non-rigid spectral correspondence of triangle meshes. *International Journal on Shape Modeling* (via invitation to Special Issue of SMI 2006).

Kawakami, R., Tan, R. T., & Ikeuchi, K. (2005). Consistent surface color for texturing large objects in outdoor scenes. In *Proceedings of the international conference on computer vision (ICCV2005)* (Vol. 2, pp. 1200–1207).

Kurazume, R., Nishino, K., Zhang, Z., & Ikeuchi, K. (2002). Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute. In *Proceedings of fifth Asian conference on computer vision (ACCV2002)* (Vol. 1, pp. 99–106).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*, 431–441.

Masuda, T., Hirota, Y., Nishino, K., & Ikeuchi, K. (2005). Simultaneous determination of registration and deformation parameters among 3D range images. In *Proceedings of 5th international conference on 3-D digital imaging and modeling (3DIM2005)* (pp. 369–376).

Matsui, K., Ono, S., & Ikeuchi, K. (2005). The climbing sensor: 3-D modeling of a narrow and vertically stalky space by using spatio-temporal range image. In *International conference on intelligent robots and systems (IROS2005)*.

Miller, R., & Amidi, O. (1998). 3-D site mapping with the CMU autonomous helicopter. In *The 5th international conference on intelligent autonomous systems*.

Miyazaki, D., Oishi, T., Nishikawa, T., Sagawa, R., Nishino, K., Tomomatsu, T., Yakase, Y., & Ikeuchi, K. (2005). The Great Buddha project: Modelling cultural heritage through observation. In *Proceedings of the 6th international conference on virtual systems and multimedia (VSMM2000)* (pp. 138–145).

Moravec, H. P. (1977). Towards automatic visual obstacle avoidance. In *Proceedings of 5th international joint conference on artificial intelligence* (p. 584).

Nishino, K., & Ikeuchi, K. (2002). Robust simultaneous registration of multiple range images. In *Proceedings of 5th Asian conference on computer vision (ACCV2002)* (pp. 454–461).

Oishi, T., Sagawa, R., Nakazawa, A., Kurazume, R., & Ikeuchi, K. (2003). Parallel alignment of a large number of range images. In *Proceedings of 4th international conference on 3-D digital imaging and modeling (3DIM2004)* (pp. 195–202).

Oishi, T., Nakazawa, A., Kurazume, R., & Ikeuchi, K. (2005). Fast simultaneous alignment of multiple range images using index images. In *Proceedings of 5th international conference on 3-D digital imaging and modeling (3DIM2005)* (pp. 476–483).

Polak, E. (1971). *Computational methods in optimization*. New York: Academic Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C*. Cambridge: Cambridge University Press.

Rusinkiewicz, S., & Levoy, M. (2001). Efficient variant of the ICP algorithm. In *Proceedings of the 3rd international conference on 3-D digital imaging and modeling (3DIM2001)* (pp. 145–152).

Smith, S. M., & Brady, M. (1997). SUSAN—a new approach to low level image processing. *International Journal of Computer Vision*, *23*(1), 45–78.

Stoer, J., & Bulirsh, R. (1980). *Introduction to numerical analysis*. New York: Springer.

Szeliski, R., & Lavallée, S. (1996). Matching 3-D anatomical surfaces with non-rigid deformations using Octree-splines. *International Journal of Computer Vision*, *18*(2), 171–186.

Thrun, S., Diel, M., & Haehnel, D. (2003). Scan alignment and 3-D surface modeling with a helicopter platform. In *Proceedings of the 4th international conference on field and service robotics*.

Wheeler, M. D. (1996). *Automatic modeling and localization for object recognition*. Ph.D. thesis, School of Computer Science, Carnegie Mellon University.

Wheeler, M. D., & Ikeuchi, K. (1995). Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(3), 252–265.

Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, *13*, 119–152.