High-resolution Shape Reconstruction from Multiple Range Images based on Simultaneous Estimation of Surface and Motion

Yoshihiro Watanabe, Takashi Komuro and Masatoshi Ishikawa
Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

Yoshihiro_Watanabe@ipc.i.u-tokyo.ac.jp

Abstract

Recognition of dynamic scenes based on shape information could be useful for various applications. In this study, we aimed at improving the resolution of three-dimensional (3D) data obtained from moving targets. We present a simple clean and robust method that jointly estimates motion parameters and a high-resolution 3D shape. Experimental results are provided to illustrate the performance of the proposed algorithm.

1. Introduction

3D shape acquisition in dynamic scenes is an important task for various applications, including robotics, inspection, man-machine interfaces, and so on. The quality of those applications mainly depends on the sensing speed and resolution. Also, naturally, the sensing system is expected to observe moving objects in real-time.

In related work, a new system capable of 955-fps shape acquisition of a high-speed moving/deforming object has been developed [19]. This performance level was achieved by integrating high-frame-rate imaging, structured-light-based triangulation, and high-speed parallel image processing. Although the speed is sufficiently high with this approach, the resolution is limited. In observing a moving object, it is not possible to use scanning of the reference light source. Therefore, the developed system must obtain the shape from a single image. In this case, the sensing method needs an operation for matching the reference projected points and observed points, and this constraint prevents an unbounded increase in the number of reference points. In addition, in order to detect the projected pattern with high accuracy in the image plane, the pattern density should be designed by taking account of the image sensor resolution.

These problems occur not just with the principle of structured-light triangulation. Another technique for shape acquisition is the time-of-flight method. There is a commercial product that obtains 3D shapes at video frame rates [1]. However, with this approach, too, the resolution improvement is limited because of the measurement accuracy. Another technique is the so-called shape from motion [11], which can acquire a 3D shape from video images taken by a single camera. The problem is that the obtained 3D points are restricted to a distinctive textured area, and therefore, the resolution is not so high. On the other hand, a technique known as shape from shading [12] allows high-resolution shape acquisition from a single image, but some knowledge of the experimental conditions is necessary in advance, including the light source, the reflectance of the material, and so on. This technique can only be used in rather narrow special cases compared with the other methods mentioned above.

The above discussion suggests that it is difficult, in terms of the general principles of 3D shape acquisition, to improve the resolution boundlessly. Therefore, these is a strong demand for a technology that can reconstruct a high-resolution shape after acquiring the sensing data.

We realized high-resolution shape reconstruction by integrating multiple time-sequential range images. The assumed situation is that a fixed sensing system observes a moving rigid target, or a moving sensing system observes a static target. The changes of relative positions in the sensing system and the target allow us to obtain range images which sample different points on the target surface at different times. These range images are expected to provide an effective approach to shape reconstruction, where the obtained resolution exceeds the limitations of the sensing system used. This task requires two processes: registration of range images and shape reconstruction. The two are not independent. The motion estimation for the registration needs continuous surface information, and the shape reconstruction needs the aligned data set. This mutual dependence between the two processes makes high-resolution shape reconstruction difficult.

This paper is organized as follows. Section 2 describes the background of the task we focused on. Section 3 explains technical requirements for this task, and the approach taken to achieve them. Section 4 details the proposed method for high-resolution shape reconstruction. Section 5 presents experiments using various shape data. Section 6 concludes with a discussion of further extensions to the method.

2. Background

Registration and surface reconstruction of range images are normally used in completely different applications or are achieved separately in related tasks, because the purpose of these operations is not the improvement of resolution. A well-known technique for the registration of range images is the Iterative Closest Point (ICP) method [4]. This method aligns two range images by minimizing the distance of corresponding 3D point pairs. The main applications have been model matching between range images [21] and range-image mosaicing from large target objects [16]. Also, the alignment of time-sequential range images obtained by structured-light triangulation utilized this method [18].

However, when multiple range images are aligned, ICP causes errors because the method works only for two range images. One method focusing on a solution to this problem is to estimate multiple motions simultaneously [15]. Also the conventional approach for this registration has mainly used high-resolution range images obtained by a scanning-based sensing system. It is important to develop other methods assuming that the input data is low resolution. One example achieved the registration of time-sequential low-resolution multiple range images [20].

After registration, surface reconstruction is applied to the aligned point set. In general, however, the original information in a real image is reduced by sampling so that the registration accuracy is considered to be insufficient for reconstructing the original surface. In addition, in the case of 3D shape acquisition, the sampling intervals are normally nonuniform on the surface. The problem of the registration accuracy is considered to be critical for this situation, where the input data is only low-resolution range images.

In related work using two-dimensional (2D) real images, the so-called super resolution technique has been advocated for a long time [13]. Super resolution is a technique for reconstructing a high-resolution 2D image from multiple low-resolution 2D images. It has been reported that it is effective to regard super resolution as an estimation problem, where registration and image restoration interact with each other [10]. The motivation of our work is related to this approach. However, the observational model and the degrees of freedom of motion that we have to use in the 3D task are totally different, and thus, a new method is required.

3. Assumptions

3.1. Requirements for the task

This section describes the problems and the key approaches employed to reconstruct a high-resolution shape from multiple range images. The framework is shown in Figure 1. The illustration shows that the technique we focus on here has two processes: motion estimation for registration and shape reconstruction.
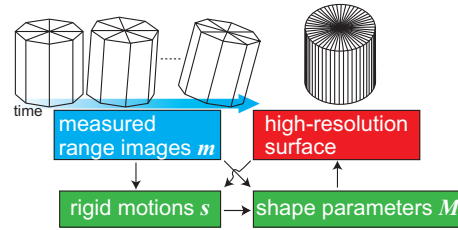

Figure 1. The process of high-resolution shape reconstruction. The input data is multiple range images. The images sample different points on the target surface and have different coordinate systems. Rigid motions and the values representing the continuous surface are mutually dependent.

As shown at the top of Figure 1, the targets captured in the low-resolution range images are on different coordinate systems from the base coordinate system, which the high-resolution reconstructed target depends on. Motion estimation obtains the motions $(s_1, s_2, \ldots, s_N)$ that align the 3D point sets in N range images onto the base coordinate system. Motion estimation using only observed images has an accuracy limit, which is caused by image degradation involving aliasing and measurement noise. Therefore, the estimation essentially needs two kinds of inputs: the measured range images $(m_1, m_2, \ldots, m_N)$ and the ideal continuous target surface in the base coordinate system, which is represented by the parameter M. In a similar way to the motion estimation, the shape reconstruction, which is the estimation of M, needs aligned range images represented by the parameters $(m_1, m_2, \ldots, m_N)$ and $(s_1, s_2, \ldots, s_N)$.

Based on this discussion, these two processes, motion estimation and shape reconstruction, are mutually dependent, necessitating the simultaneous estimation of all parameters.

3.2. Problem formulation

Our task is formulated as the following minimization problem:

$$\min_{M, s} d(\xi(M, u), \tilde{m}(s)) \tag{1}$$

Here, the function d outputs the distance between the reconstructed shape and the range images. ξ is an example representation of the reconstructed continuous shape surface. M are the base parameters to represent it, and u are variable values. $\tilde{m}(s)$ is the aligned point set. These parameters are related by the following equation.

$$\tilde{m}(s) = \{R(s_f)m_{f,i} + T(s_f) | f = 1, \ldots, N, i = 1, \ldots, L_f\} \tag{2}$$

The i-th 3D point of the f-frame range image is set as $m_{f,i}$. The f-frame rigid motion is expressed as $R(s_f)/T(s_f)$, where s_f are the motion parameters. There are six degrees of freedom.

The minimization problem written as (1) asks technical questions about distance and continuous shape representa-

tion. First, in this paper, the distance is simply defined as follows:

$$d(\boldsymbol{\xi}(\boldsymbol{M}, \boldsymbol{u}), \tilde{\boldsymbol{m}}(\boldsymbol{s})) = \sum_{f=1}^{N} \sum_{i=1}^{L_f} |\boldsymbol{A}_{f,i}(\boldsymbol{\xi}) - \tilde{\boldsymbol{m}}_{f,i}(\boldsymbol{s}_f)|^2. \quad (3)$$

Here, $\boldsymbol{A}_{f,i}$ is a function to pick up the point on the continuous target surface that corresponds to the aligned point $\tilde{\boldsymbol{m}}_{f,i}(\boldsymbol{s}_f)$. This function $\boldsymbol{A}$ represents the observation process involving downsampling, shown in Figure 2. This type of observation model is introduced in the registration of range images for the accuracy improvement [6]. Unlike the super-resolution of a 2D image, the point spread function due to the lens is not included in this observation model. However, the sampling intervals are nonuniform on the observed plane. Downsampling normally depends on the sensing system. As an example of the observational model, in the case of time-of-flight method, the intersection of the corresponding projected rays defines the point $\boldsymbol{A}_{f,i}(\boldsymbol{\xi})$.
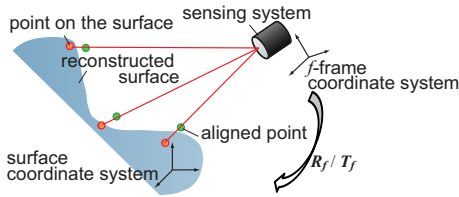

Figure 2. Distance between the surface and the aligned point set. We define the distance based on the observational model.

The other problem is the representation of the continuous surface $\boldsymbol{\xi}$. Three representations have mainly been used: a mesh-based surface, a parametric surface, and an implicit surface.

The mesh representation is the simplest one. The surface is represented as a set of planes connecting the element points directly. In this case, the parameter $\boldsymbol{M}$ is the 3D point set. In order to represent a shape with high resolution, however, it should be large, thus making it too difficult to solve the estimation.

The parametric surface includes a bezier surface, a nurbs surface, and so on. Another technique considered to be effective is surface interpolation by a sinc function, because it involves band limitation of the space-domain frequency of the surface. In high-resolution shape reconstruction from range images, the overfitting problem, where the estimated shape is fitted to a misaligned point set, is an important issue. A novel shape representation method, such as the sinc-interpolation approach, can reject the noisy alignment, allowing convergence to an optimized surface. In the parametric surface approach, the parameter $\boldsymbol{M}$ is the set of base points, whose number is drastically decreased compared with the mesh representation. However, the disadvantage with this approach is the need to divide the whole shape into parts to realize high degree of freedom for the surface, including a closed surface like a sphere. Also, the

method based on sinc-interpolation causes overshoot and undershoot problems if the set of base points is not adequately arranged.

The third representation is based on an implicit surface. This technique represents the surface by using a scalar function $f(\boldsymbol{m})$ at 3D position $\boldsymbol{m}$. This function is set under the condition that, if the point $\boldsymbol{m}$ is on the surface $\boldsymbol{\xi}$, the function satisfies the following equation:

$$f(\boldsymbol{m}) = 0. \quad (4)$$

Recently, various implicit surface techniques have been proposed, for example, those based on a Radial Basis Function (RBF) [5], Multi-level Partition of Unity [17], and Moving Least Squares [3]. The implicit surface can represent various surfaces and can control the local smoothness on the surface easily without breaking its basic formulation.

These discussions suggest that the implicit surface is an effective shape representation because it can realize the requirements of our reverse problem, such as small variable estimated parameters, high-degree of freedom of the surface, and easy control of local surface changes. In this paper, therefore, we employed the method using RBF [5]. This function has the form

$$f(\boldsymbol{m}) = p(\boldsymbol{m}) + \sum_{i=1}^{N_B} \lambda_i \phi(|\boldsymbol{m} - \boldsymbol{m}_i|) \quad (5)$$

where p is a polynomial function, and ϕ is a basic function, which includes the biharmonic spline $\phi(r) = r$, the thin-plate spline $\phi(r) = r^2 log(r)$, and so on. In this paper, we used the biharmonic spline.

4. High-resolution Shape Reconstruction from Multiple Range Images

4.1. Overview

As mentioned in the previous section, the shape $\boldsymbol{M}$ and motion $\boldsymbol{s}$ must be estimated at once. However, it is difficult to achieve this with high estimation accuracy. Therefore, we propose the following approximated simultaneous estimation:

1. Get the initialized shape parameters $\boldsymbol{M}^0$.
2. Estimate the motions $\boldsymbol{s}^i$ by using the estimated shape parameters $\boldsymbol{M}^{i-1}$.
3. Estimate the shape parameters $\boldsymbol{M}^i$ from the point set aligned by the estimated motions $\boldsymbol{s}^i$.
4. Repeat steps 2 and 3 until convergence.

This method estimates the motions and shape alternately. By repeating the cycle, we can obtain a converged solution. However, in the beginning cycles, the input point set is not aligned correctly so that the input data in the shape reconstruction is regarded as being noisy. On the other hand, in the last cycles, where the point set is expected to be aligned

correctly, the input data in the reconstruction is considered to have less noise. This suggests that the conditions to the data in the shape reconstruction should be changed as the iterative estimation process proceeds. The details are described in the following sections.

4.2. Motion estimation with implicit surface

In the motion estimation, the target shape is fixed. The task is to estimate the motions that bring the point set closer to the base coordinate system. However, both the input point set and the target surface are assumed to have errors caused by estimation errors and measurement noise. These problems cause a situation where some points cannot be placed close to the surface. Therefore, our motion estimation requires a robust estimation technique. The proposed method is based on RANdom SAmple Consensus (RANSAC) [9]. With this method, N motions for N frames can be estimated individually. The details of the first part of each motion estimation are listed below.

1. Randomly select l_r points from the f-frame data set $\boldsymbol{m}_f$.

2. Decide the correspondences $\boldsymbol{A}_{f,i}(\boldsymbol{\xi})$ to the selected points $\boldsymbol{x}_{f,i}$ based on the observational model.

3. Estimate the motion minimizing the sum of the distances of the points $\boldsymbol{x}_f$ and $\boldsymbol{A}_f(\boldsymbol{\xi})$.

4. Calculate the distance between the surface and all moved points $\tilde{\boldsymbol{m}}_f$.

5. Count the number of points whose distance is within a threshold.

6. Repeat steps 1 to 5 and choose the solution with the largest score.

In step 2, the correspondences are decided based on the observational model. In the time-of-flight case, the intersection of the surface and the projected ray passing through the point $\boldsymbol{x}_{f,i}$ is considered to be the correspondence $\boldsymbol{A}_{f,i}(\boldsymbol{\xi})$. This is shown in Figure 2. The observational model can be tuned according to the sensor system.

The minimization problem in step 3 is formulated as:

$$\min_{\hat{\boldsymbol{s}}_f} \sum_{i=1}^{l_r} |\boldsymbol{A}_{f,i}(\boldsymbol{\xi}) - \boldsymbol{R}(\hat{\boldsymbol{s}}_f)\boldsymbol{x}_{f,i} - \boldsymbol{T}(\hat{\boldsymbol{s}}_f)|^2. \quad (6)$$

Several solutions have been proposed for this type of rigid motion estimation [8].

The above steps provide the temporal solution for f-frame motion $\hat{\boldsymbol{s}}_f$. Using this motion $\hat{\boldsymbol{s}}_f$, the f-frame point set can be divided into inliers and outliers. The operation is based on the distance to the surface. After these operations, optimal estimation is applied. Optimal estimation has two steps. First, motion is estimated in similar ways in steps 2 and 3 by using only L'_f inlier points which are obtained

by the above steps. After this estimation, estimation based on the following minimization problem is applied as a final adjustment:

$$\min_{\boldsymbol{s}_f} \sum_{i \in L'_f \ inliers} f(\tilde{\boldsymbol{m}}_{f,i}(\boldsymbol{s}_f))^2. \quad (7)$$

This estimation is solved by using a non-linear estimation method. As the initial values, we used the parameters estimated by the first step of the optimal estimation. This estimation allows us to obtain the optimized solutions for the shape reconstruction based on an implicit surface.

4.3. Shape reconstruction

The shape surface $\boldsymbol{\xi}$ is estimated from the aligned point set $\tilde{m}$. The shape reconstruction problem is formulated as:

$$\min_{\boldsymbol{M}} \sum_{f=1}^{N} \sum_{i=1}^{L_f} |\boldsymbol{A}_{f,i}(\boldsymbol{\xi}) - \tilde{\boldsymbol{m}}_{f,i}(\boldsymbol{s}_f)|^2 + g(\boldsymbol{\xi}). \quad (8)$$

We need to estimate the parameters $\boldsymbol{M}$ representing the surface of the target shape. The function g is the constraint for the surface. Shape reconstruction based on the implicit surface approximately replaces this problem with:

$$\min_{f} \|f\|^2 + \sum_{f=1}^{N} \sum_{i=1}^{L_f+\alpha} \frac{1}{\rho_f} (f(\tilde{\boldsymbol{m}}_{f,i}(\boldsymbol{s}_f)) - f_{f,i})^2. \quad (9)$$

The constraints part g is replaced by the first term, which is described by:

$$\|f\|^2 = \int_{\mathbb{R}^3} \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial y^2} \right)^2 + \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial z^2} \right)^2 +$$
$$2 \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial x \partial y} \right)^2 + 2 \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial x \partial z} \right)^2 + 2 \left(\frac{\partial^2 f(\boldsymbol{m})}{\partial y \partial z} \right)^2 d\boldsymbol{m}. \quad (10)$$

The parameter $f_{f,i}$ in (9) is a given parameter to define the scalar function $f(\boldsymbol{m})$. The second term of (9) has additional given points that are not included in the aligned point set and are defined to have non-zero values to solve this problem correctly.

Also, the problem (9) has new parameter ρ_f. Considering that the estimated motions have errors, exact fitting causes the problem of overfitting. Thus, the weight of each frame's point set depends on the accuracy of the motion estimation. The implicit surface based on the RBF function allows this approach to be realized simply by using this parameter ρ_f. The parameter ρ_f controls the weight of the input data. If the parameter ρ_f is small, the reconstructed shape surface passes the f-frame point set. If the parameter ρ_f is large, the f-frame point set is regarded as noisy data, and smooth interpolation is achieved.

It has been shown that the solution of (9) is obtained by solving the linear system [7], given by:

$$\begin{pmatrix} A - diag(\boldsymbol{\rho}) & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} \\ 0 \end{pmatrix} \quad (11)$$

where $A_{p,q} = \phi(|\boldsymbol{m}_p - \boldsymbol{m}_q|)$, and $P_{p,q} = p_q(\boldsymbol{m}_p)$. Here, p_q is the q-th value of the polynomial function $p(\boldsymbol{m})$. Also $diag(\boldsymbol{\rho})$ is the diagonal matrix with the elements $\boldsymbol{\rho}$. The point value with a single suffix means $\boldsymbol{m}_p = \boldsymbol{m}_{i,j}$, where $p = \sum_{f=2}^{i} L_{f-1} + j$.

5. Experiments

5.1. Stanford bunny
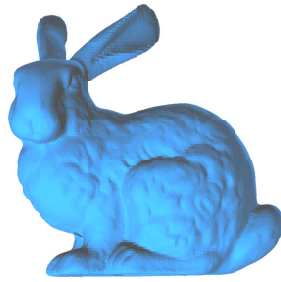
Figure 3. Stanford bunny.

We evaluated the proposed method with some example shapes. The first one was the Stanford bunny. The 3D data of the Stanford bunny was obtained from [2]. The rendered shape is shown in Figure 3. The number of mesh triangles constructing the Stanford bunny was 69,451. The size was $78 \times 77 \times 60$.

The intersecting points with lines projected in a radial pattern were used as the measured points. At each frame, the target was moved by rotation and translation. The rotation axis was $(1, 1, 0)$, the rotation angle was 0.03 rad, and the translation was $(0.2, 0.2, -0.15)$. The total number of frames was 30. The total number of 3D points was 1,561. The coordinate system of the 15-th frame (base frame) was set as a base coordinate system and was fixed. The motions of point sets in other frames, which approached the base coordinate system, were estimated. Figure 4 (a) shows a single range image in the base frame. The initial shape surface is shown in Figure 4 (b). The surface was estimated from the point set in the base frame. The initial surface was estimated with an implicit surface using the RBF function, as described above. The mesh for surface rendering was extracted by the technique called Marching Cubes [14].

In the calculations, the number of selected points in RANSAC motion estimation l_r was 8, and the number of iterations was 30. The number of iterations for motion estimation and shape reconstruction was 5. Also, in this experiment, the constraints $\boldsymbol{\rho}$ in shape reconstruction were simply set. Since input points in the base frame are definitely on the surface, zero values were provided. Also for
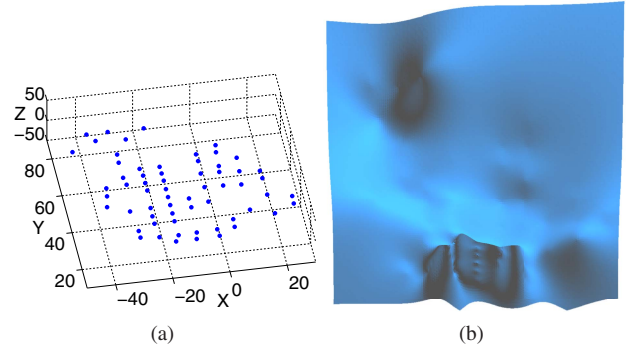

Figure 4. A single range image of the Stanford bunny. (a) A single range image in the base coordinate system. (b) An initial shape surface was estimated from a single range image in the base frame.
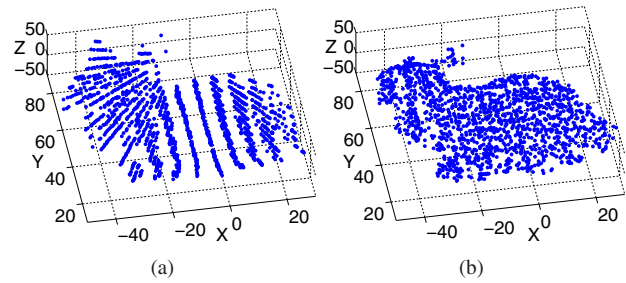

Figure 5. The registration of the Stanford bunny. (a) The point set before registration. There are point sets for 30 frames. (b) An aligned point set is drawn. The proposed method aligned 30 frame point sets correctly.
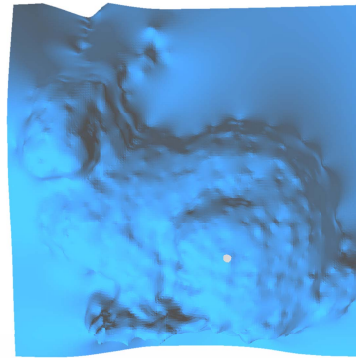

Figure 6. The shape reconstruction of the Stanford bunny. The reconstructed surface was rendered.

the other points, the same balance parameters were globally provided. In addition, the parameters were changed so as to become smaller with every iteration. This is because the accuracy of the surface was expected to improve with each iteration.

Figure 5 (a) shows all point sets for 30 frames before registration. The aligned point set after 5 iterations is shown in Figure 5 (b). Compared with Figure 5 (a), the point set was aligned correctly. The reconstructed shape is shown in Figure 6. The face and foot of the bunny, which could not

be found in the initial shape surface shown in Figure 4 (b), newly appeared in the final shape.
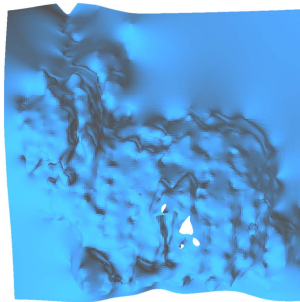

Figure 7. The estimated shape of the Stanford bunny in the comparative method.

In addition, we compared the results with another method. In the comparative method, registration and shape reconstruction were carried out separately, as in the conventional approaches. The registration was achieved based on ICP. ICP was iteratively applied to two range images, the base frame and other frames. The correspondence pair in ICP was set as the closest points in two point sets. After that, the surface was interpolated with an implicit surface using the RBF function. The interpolated surface is shown in Figure 7.

The surface accuracy in the proposed method and the comparative one was evaluated based on the distance of the points on the estimated surface and the true one. The points at the same (x, y) position were selected as the correspondence pair. As a result, the average accuracy per evaluated point using the comparative method was 1.11. The accuracy using the proposed method was 0.40. The accuracy was improved with every iteration. The changes of the values were (0.89, 0.78, 0.71, 0.41, 0.40).
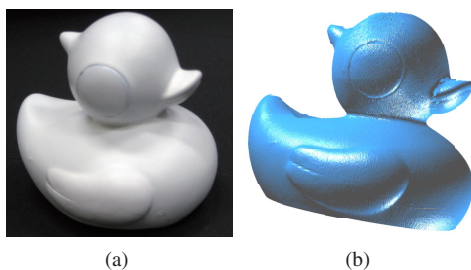
5.2. Bird doll

Figure 8. Bird doll. (a) A picture of the object. (b) The shape of the bird doll. The shape was measured with the laser scanner.

The second example was a bird doll. A picture of this doll is shown in Figure 8 (a). The detailed shape of this object was obtained by a 3D laser scanner (Roland LPX-600). The scanned shape is shown in Figure 8 (b). The number of mesh triangles constructing this bird doll was 288,300. The size was 68 mm × 65 mm × 58 mm. By using
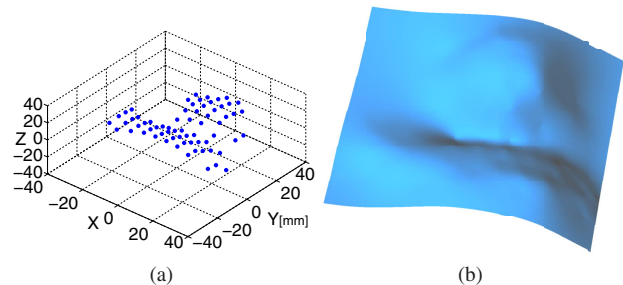

Figure 9. A single range image of the bird doll. (a) A single range image in the base coordinate system. (b) Initial shape surface was estimated from a single range image in the base frame.
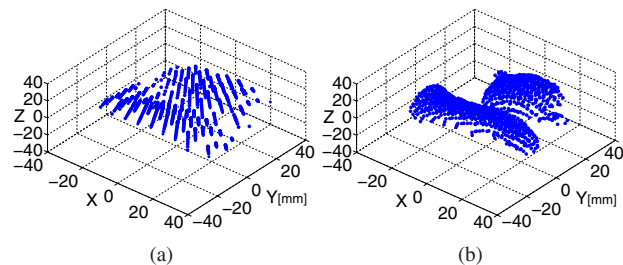

Figure 10. The registration of the bird doll. (a) The point set before registration. There are point sets of 20 frames. (b) An aligned point set was drawn. The proposed method aligned 20 frame point sets correctly.
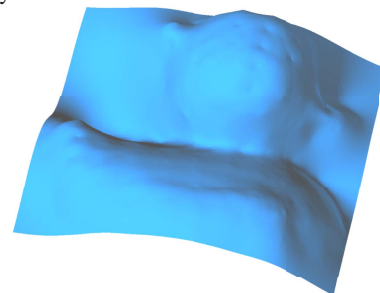

Figure 11. The shape reconstruction of the bird doll. The reconstructed shape surface was rendered.

this data, the input range images for this experiment were created with the same method as in the previous section.

In this experiment, the rotation axis was set as $(0, 1, 0)$, the rotation angle per a frame was 0.05rad, the and translation per a frame was $(0.5 \text{ mm}, 0.5 \text{ mm}, -0.3 \text{ mm})$. The total number of frames was 20, and the total number of 3D points was 1,269. The base coordinate system was the 10-th frame. Figure 9 (a) shows a single range image in the base frame. The initial shape surface is shown in Figure 9 (b).

The parameter settings were the same as in the experiment described above. Figure 10 (a) shows all point sets before registration. The aligned point set after 5 iterations is shown in Figure 10 (b). The reconstructed shape is shown in Figure 11. The distinctive changes from the initial shape surface can be found in the beak, the jut on the back of the

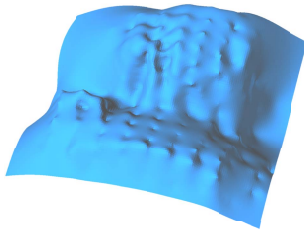head, and the tail, which cannot be obtained from a single frame.

Figure 12. The estimated shape of the bird doll in the comparative method.

The evaluation was performed in a similar way to the previous experiment. The estimated shape in the comparative method is shown in Figure 12. The accuracy per evaluated point using the comparative method was 1.64 mm. The accuracy using the proposed method was 0.25 mm. The changes of the accuracy at every cycle were (0.44, 0.41, 0.40, 0.26, 0.25).
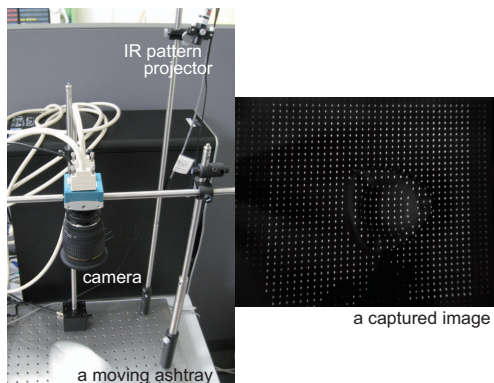
5.3. Ashtray

Figure 13. Picture of sensing system and one of the captured images. The each spot in an image was converted to a 3D point based on triangulation.

In this experiment, we used the data obtained by a structured-light sensing system. The system projected a 1,000-spot pattern (StockYale pattern projector) onto the target and triangulated at each spot based on the position observed in the captured image plane. The camera (Mikrotron Eosens) captured the scene at a frame rate of 500 fps. The resolution was $1,280 \times 1,024$. A picture of the system is shown in Figure 13. The target was moved within the measured area. The captured image during the measurement is also shown in Figure 13. The measured target was an ashtray (Figure 14). The size was 107 mm $\times$ 108 mm $\times$ 27 mm.

In this experiment, the rotation axis was $(0, 0, 1)$, the rotation angle per a frame was 0.02 rad, and the translation per a frame was $(-0.8 \text{ mm}, -1.2 \text{ mm}, 0.0 \text{ mm})$, which were estimated parameters. The total number of frames was 20, and
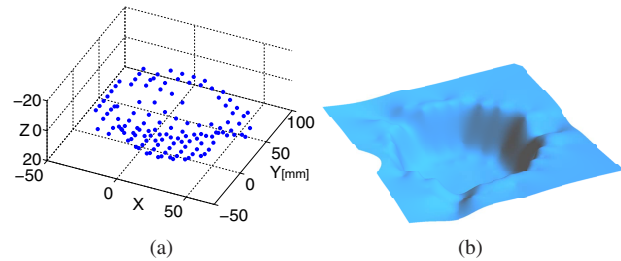

Figure 14. Ashtray.

Figure 15. A single range image of the ashtray. (a) A single range image in the base coordinate system. (b) The initial shape surface was estimated from a single range image in the base frame.
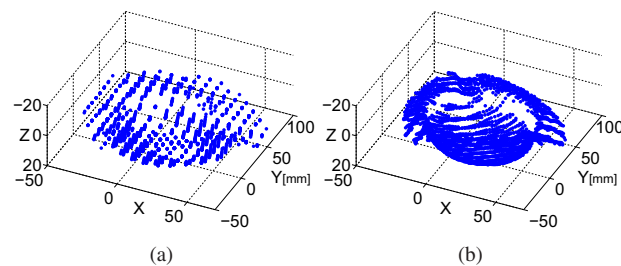

Figure 16. The registration of the ashtray. (a) The point set before registration. There are point sets of 20 frames. (b) An aligned point set was drawn. The proposed method aligned 20 frame point sets correctly.
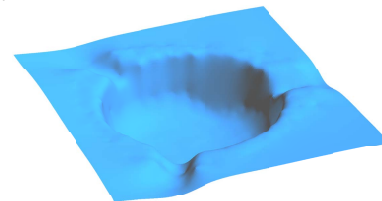

Figure 17. The shape reconstruction of the ashtray. The reconstructed shape surface is rendered.

the total number of 3D points was 2,498. The base coordinate system was the 10-th frame. Figure 15 (a) shows a single range image in the base frame. The initial shape surface is shown in Figure 15 (b). The measured noise was considered to be larger than the experiments mentioned above.

The parameter settings were the same as in the previous experiments. Figure 16 (a) shows all point sets before registration. The aligned point set after 5 iterations is shown in Figure 16 (b). The reconstructed shape is shown in Figure 17. The corner shape at the top is cleanly represented in

the reconstructed shape. Evaluation of this data is left as the topic of future work because the orientations of the true surface and the measured one in the base frame were difficult to match.

6. Conclusion

High-resolution shape reconstruction from acquired range images is a highly promising technique for various applications, regardless of sensing principle. This task involves two processes: registration and surface reconstruction. Those processes are essentially interrelated problems, and performing registration using only the acquired range images causes errors because of the image degradation involving aliasing and measured noise. Also, the shape surface representation needs to have a continuous form and a high degree of freedom with a compact data set. In addition, local smoothness control is an important issue for surface reconstruction.

Based on these requirements, in this paper, we present an algorithm for estimating the motions and surface jointly. The method realized shape reconstruction based on an implicit surface obtained using an RBF function. We demonstrated the potential of this approach with three kinds of real shapes. The experiments were successfully achieved. The obtained results show that we can see details on the surface that cannot be found in a single range image.

As future work, we think it is important to develop a more effective technique to control the parameters for constraints in the process of shape reconstruction. Also, although the described experiment observes the smaller object than measured range, it is important to reconstruct a large object. This problem could be solved by enlarging the base shape based on the estimated results. In addition, an improvement in calculation speed, possibly by using a parallel processing module, is required in order to combine this approach with a real-time sensing system.

References

[1] MESA imaging AG, http://www.mesa-imaging.ch/.

[2] The stanford 3D scanning repository, http://graphics.stanford.edu/data/3dscanrep/.

[3] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva. Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(1):3–15, 2003.

[4] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[5] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 67–76, 2001.

[6] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, 1996.

[7] H. Q. Dinh and G. Turk. Reconstructing surfaces using anisotropic basis functions. In *In the Proceedings of International Conference on Computer Vision*, pages 606–613, 2001.

[8] D. W. Eggert, A. Lorusso, and R. B. Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9(5):272–290, 1997.

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

[10] R. C. Hardie, K. J. Barnard, E. E. Armstrong, and S. T. Branch. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6:1621–1633, 1997.

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

[12] B. K. Horn and M. J. Brooks. *Shape from Shading*. MIT Press, 1989.

[13] M. G. Kang and S. Chaudhuri. Super-resolution image reconstruction. *IEEE Signal Processing Magazine*, 20(3):19–20, 2003.

[14] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

[15] P. J. Neugebauer. Geometrical cloning of 3D objects via simultaneous registration of multiple range images. In *Proceedings of the 1997 International Conference on Shape Modeling and Applications*, pages 130–139, 1997.

[16] K. Nishino and K. Ikeuchi. Robust simultaneous registration of multiple range images. In *Proceedings of the 5th Asian Conference on Computer Vision*, pages 23–25, 2002.

[17] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H.-P. Seidel. Multi-level partition of unity implicits. *ACM Trans. Graph.*, 22(3):463–470, 2003.

[18] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. *ACM Transactions on Graphics*, 21(3):438–446, 2002.

[19] Y. Watanabe, T. Komuro, and M. Ishikawa. 955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis. In *Proceedings of 2007 IEEE International Conference on Robotics and Automation*, pages 3192–3197, 2007.

[20] Y. Watanabe, T. Komuro, and M. Ishikawa. Integration of time-sequential range images for reconstruction of a high-resolution 3D shape. In *Proceedings of 19th International Conference on Pattern Recognition*, 2008.

[21] J. Ziegler, K. Nickel, and R. Stiefelhagen. Tracking of the articulated upper body on multi-view stereo image sequences. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2006.