Integration of Time-sequential Range Images for Reconstruction of a High-resolution 3D Shape

Yoshihiro Watanabe   Takashi Komuro   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

The recognition of dynamic scenes using 3D shapes could provide useful approaches for various applications. However, the conventional 3D-shape sensing systems dedicated for such scenes have had problems in spatial resolution, though they have achieved high sampling rate in temporal domain. In order to solve this limits, we present a method that integrates time-sequential partial range images capturing moving targets to reconstruct a high-resolution range image. In the proposed method, multiple range images are set in the same coordinate system based on multi-frame simultaneous alignment. This paper also demonstrates the performance of the proposed method using some example rigid bodies.

1. Introduction

The task we address in this paper is to realize high-speed and high-resolution 3D-shape sensing for moving objects in real-time. This could be promising technology in applications where 3D recognition of dynamic scenes is required, such as robotic control; surgical support, for example, observation of a beating heart [1]; rapid visual inspection of products; automotive applications, for example, checking road surfaces; and human-machine interaction with high accuracy and flexibility.

In order to realize such 3D-shape sensing, there are some challenges. In this paper, we focus on the reconstruction of a high-resolution range image from low-resolution, time-sequential partial data obtained by observing moving targets. The problem of this time-sequential 3D-shape integration is divided into three tasks: (A) segmentation of a scene into rigid bodies,

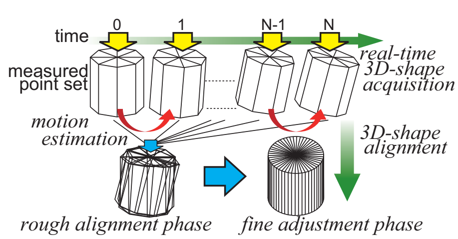


Figure 1. Integration of time-sequential range images.

(B) alignment of range images by estimating motion of each rigid body, and (C) reconstruction of curved surfaces. The concept is shown in Figure 1. This paper presents a method for (B).

2. Real-time 3D-shape Sensing of Moving/Deforming Objects

Conventional systems are mainly employed in situations where the relationships among the objects being measured and observation devices are completely known, by ensuring stable or controllable conditions. In applications where dynamic scenes are observed, this limitation has become a critical problem.

To overcome this problem, we have developed a new 3D-shape sensing system[6]. The system acquires a 3D shape from a single image based on structured-light-projected triangulation, allowing it to observe a high-speed moving rigid body or a deforming or vibrating non-rigid body in real-time. It achieved 955-fps throughput and 4.5-ms latency based on its high-frame-rate imaging and high-speed image processing.

Also there is another example developed by Rusinkiewicz [4] that has combined the structured light method and the method aligning multiple range images. The system achieved the full-view 3D shape reconstruction for an object moved by hand in front of the system. Although its sensing speed is not high enough for real-time applications, the concept is highly relevant with our goals. This work used the Iterative Closest Point (ICP) as the data alignment method. The method has been widely employed and some extended ones have been developed according to each application [5]. However, the situation where the resolution of input range images is low have not been tried yet.

In the structured-light-based 3D-shape sensing, the improvement of the resolution is limited by two factors. One is the registration problem of matching the reference projected points and observed points in an image. The other is the detection accuracy in an image. These factors become favorable when the resolution is low and the size of an element in a pattern whose position is to be detected is large. Therefore, to overcome this

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

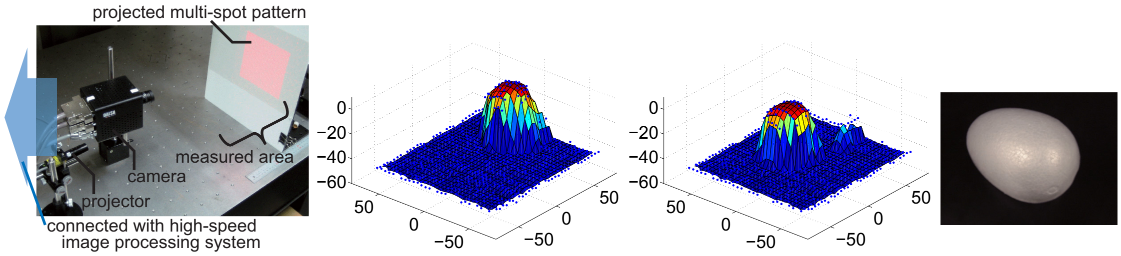


Figure 2. 955 fps 3D-shape real-time sensing system[6]. The images at the center show the 3D ball shapes extracted every 30 ms from the entire data. The right picture is the target object.

limitation, the sensing systems require a new method to reconstruct a high-resolution range image from low-resolution ones after data acquisition.

3. Strategy for Time-sequential Data Alignment

In this paper, the addressed task is motion estimation and data alignment by using time-sequential range images containing a rigid body. ICP has been effective for this type of task. This method uses two range images and sets corresponding points in those images. The motion of a rigid body is estimated by minimizing the distances between those points. The main application of this method has been model matching between range images [7] and range-image mosaicing from large target objects [3].

The basic framework of this method is expected to be useful in our focused task. However, there are two assumptions causing less-accurate alignment. First, the low-resolution 3D point sets are assumed to be inputted. Second, the multiple range images needs to be aligned. The key points of our proposed method are an estimation method based on setting of point to plane correspondence and simultaneous optimization for rigid motions of multiple range images. Also, our method aligns range images in a phased manner in order to maintain stable estimation accuracy. The method is divided into two phases: a rough alignment phase and a fine adjustment phase. Details are shown in section 4 and 5, respectively.

4. Shape Alignment using Two Images

Two-image alignment used in rough alignment is basically based on the ICP approach. We have two successive images $\boldsymbol{I}(t)$ and $\boldsymbol{I}(t+1)$. The image $\boldsymbol{I}(t)$ consists of 3D points $\{\boldsymbol{x}_i(t)|i=1,\ldots,n(t)\}$. Here $n(t)$ means the number of points in $\boldsymbol{I}(t)$. First, we need to select/generate the reference point $\boldsymbol{y}_j(t+1)$ from $\boldsymbol{I}(t+1)$ that corresponds to the target point $\boldsymbol{x}_j(t)$.

Conventional approaches have mainly searched nearest points from two images and set them as corresponding points. This has not become problem in the case using high-resolution images. However, this task has low-resolution input images and a low likelihood of obtaining 3D points at the same position on the target face, which are considered to cause low-accurate alignment.

Those discussions suggest that it is more effective for this task to generate a hypothetical reference point $\boldsymbol{y}_j(t+1)$ instead of simply using nearest measured points. Here, we employ the point-to-plane scheme. In this scheme, the point on the plane passing through the three points nearest the target point $\boldsymbol{x}_j(t)$ is set as the reference point $\boldsymbol{y}_j(t+1)$, as follows:

$$\boldsymbol{y}'_j = \frac{(1-\boldsymbol{X}'^{-1}_n\cdot\boldsymbol{x}_j)((\boldsymbol{x}'_b-\boldsymbol{x}'_a)\times(\boldsymbol{x}'_c-\boldsymbol{x}'_a))}{(\boldsymbol{X}'^{-1}_n\boldsymbol{\psi})\cdot((\boldsymbol{x}'_b-\boldsymbol{x}'_a)\times(\boldsymbol{x}'_c-\boldsymbol{x}'_a))}+\boldsymbol{x}_j \tag{1}$$

Here, $\boldsymbol{X}'_n=[\boldsymbol{x}'_a,\boldsymbol{x}'_b,\boldsymbol{x}'_c]^t$ are the three points nearest the target point $\boldsymbol{x}_j$ and $\boldsymbol{\psi}$ is $[1,1,1]^t$. The values with primes are the points at time $(t+1)$. The proposed scheme is illustrated in Figure 3. In the figure, $\boldsymbol{P}_j(t+1)$ is the plane consisting of the points $\boldsymbol{X}'_n$.

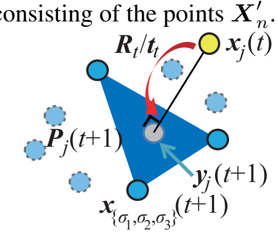


Figure 3. Corresponding points between two point sets.

Based on this scheme, the cost function for estimation of rigid motion is as follows:

$$E(t)=\sum_j\|\boldsymbol{y}_j(t+1)-\boldsymbol{R}(t)\boldsymbol{x}_j(t)-\boldsymbol{t}(t)\|^2 \tag{2}$$

The rigid motion $\boldsymbol{R}(t)/\boldsymbol{t}(t)$ at time t which minimizes this function is estimated. This means that all points on the target surface have nearest planes, and the whole 3D shape minimizes the sum of their distances. This assumption is considered to be valid except the objects containing jagged surface. It should be possible to apply this method to many situations.

The algorithm steps are as follows: (1) Select the corresponding points based on the point-to-plane scheme. (2) Estimate the rigid motion based on the cost function. (3) Apply the rigid motion. (4) If the value of the cost function is small, the calculation is finished. If not, repeat steps (1) to (3).

5. Multi-frame Time-series Simultaneous Alignment

From the estimated motions by applying the two-image shape alignment to all images, it is possible to align the multiple range images in the same relative coordinate system. However, this aligned image has cumulative motion errors, as shown in Figure 1. In this section, we describe fine adjustment which is achieved by using multiple range images and optimizing the errors simultaneously.

One conventional approach is the method of Neugebauer [2]. This method defines the sum of the distances between corresponding points in every pair of range images as the cost function and estimates the multiple motions simultaneously. However, the accuracy of this method is not much higher than that using only two images because the reference point is calculated from a set of points of a single range image.

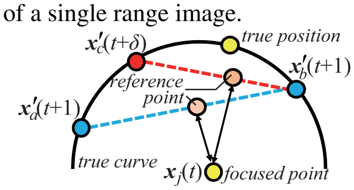


Figure 4. Selection scheme for multi-frame alignment.

In our adjustment, a reference point is generated by using multiple range images at once. Here, in order to set a reference point corresponding to the target point $\boldsymbol{x}_j(t)$, the nearest three points for the plane $[\boldsymbol{x}_{\sigma_1^j}(t+\delta_1), \boldsymbol{x}_{\sigma_2^j}(t+\delta_2), \boldsymbol{x}_{\sigma_3^j}(t+\delta_3)]$ are selected from all range images except the points in the same set in order to utilize higher resolution range images. Figure 4 shows an example. In the figure, compared to the case where only a single image is used, using multiple images to select the reference point approaches the true position, leading to improved accuracy.

With demands of this change, the cost function is also enhanced. In the two-image alignment method in section 4, only the target points are moved to the reference points. In contrast, in this multi-frame alignment, the reference points are also moved because the cost function includes all rigid motions. Based on this, the cost function E_{multi} is formulated as follows:

$$E_{multi} = \sum_t \sum_j d(\boldsymbol{P}_j(t), \boldsymbol{x}_j(t)) \tag{3}$$

$$d(\boldsymbol{P}_j(t), \boldsymbol{x}_j(t)) = \min_{u+v+w=1} \| u\hat{\boldsymbol{x}}_{\sigma_1^j}(t+\delta_1) + v\hat{\boldsymbol{x}}_{\sigma_2^j}(t+\delta_2) + w\hat{\boldsymbol{x}}_{\sigma_3^j}(t+\delta_3) - \hat{\boldsymbol{x}}_j(t) \|^2 \tag{4}$$

$$\hat{\boldsymbol{x}}(t) = \boldsymbol{R}(t)\boldsymbol{x}(t) + \boldsymbol{t}(t) \tag{5}$$

This estimation is solved as a nonlinear minimization problem to give the estimated motion parameters. The algorithm steps are performed in the same way as for the two-image alignment. Also, the image serving as the basis of the coordinate system to which all points are aligned is fixed.

6. Experiments

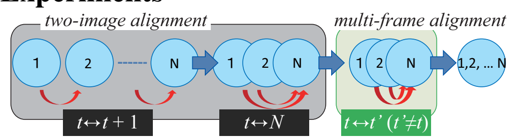
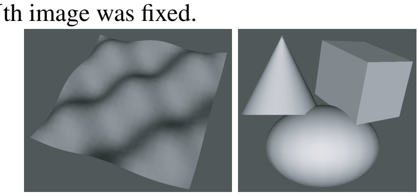


Figure 5. Calculations performed in experiments.

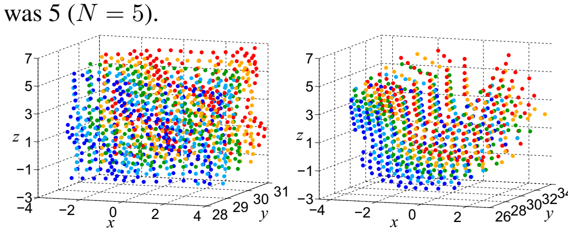
The algorithm performed in the experiments is shown in Figure 5. In the figure, each circle represents a set of 3D points and the distances between them indicate the alignment accuracy. The algorithm includes three steps. First, the two-image alignment is applied to two successive images. After applying the first estimated rigid motion, the two-image alignment is then applied to each image and the final image again. Finally, multi-frame simultaneous alignment is applied to all images. In this evaluation, N images were used and the Nth image was fixed.



(a) curved surface (b) rigid bodies

Figure 6. Rendered images.

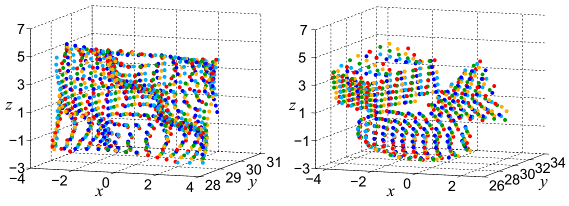
First, we show the simulated results. The used virtual objects were wavelike curved surface and connected rigid bodies. The rendered images are shown in Figure 6. They were discretized at intersecting points with lines projected from an origin in a radial pattern. In this experiments, the total number of range images was 5 ($N = 5$).



(a) curved surface (b) rigid bodies

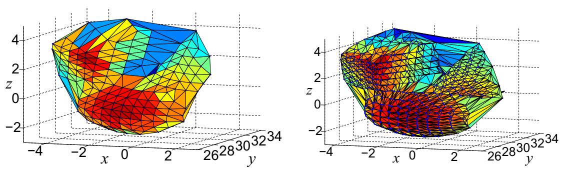
Figure 7. Initial alignment of each object.

Each initial alignment of the point set is shown in Figure 7. The points for five images are shown in different colors. The total number of 3D points in curved surface and rigid bodies were 1,122 and 821, respectively.



(a) curved surface (b) rigid bodies

Figure 8. Aligned point set of each object.



(a) a single range image (b) an aligned range image

Figure 9. Range images.

The final aligned point set is shown in Figure 8. In addition, Figure 9 shows a comparison of a single image and an integrated image in case of rigid bodies. These images were drawn by connecting the nearest points to clearly show its surface shape.

The errors in curved surface obtained after two-image alignment and multi-frame alignment were 0.12 and 0.11, respectively. Here, the error means the average distances at each point between the moved points based on motion estimation and the true points. Also, the errors in rigid bodies were 0.19 and 0.15. In addition, the errors in those objects after the two-image alignment based on conventional approach, where two nearest points were simply set as corresponding points, were 0.27 and 0.33.

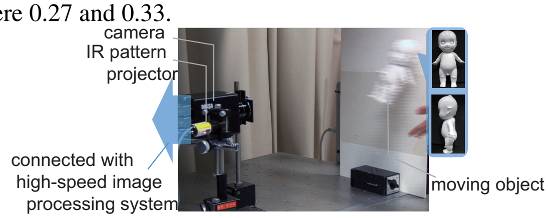
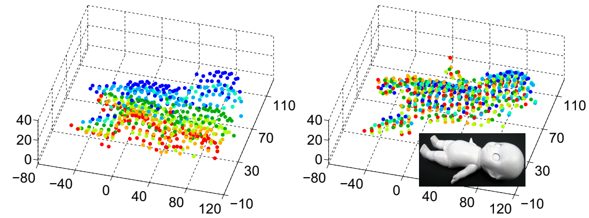


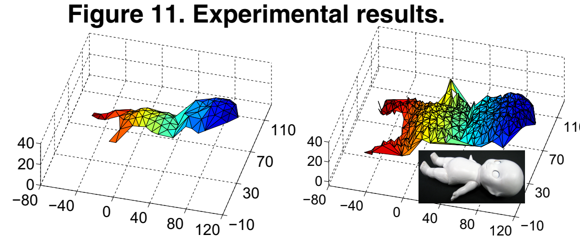
Figure 10. Experimental setup for evaluation

Next results were obtained by using data measured in the experimental environment shown in Figure 10. The used object is also shown in it, whose size was 17cm $\times$ 10cm $\times$ 6cm. Here, the developed system [6] was used. As a reference light, 33 $\times$ 33 multi-spot pattern was projected. The high sampling rate of this system is expected to realize the situation, in which all we have to do for high-resolution data acquisition is to make objects speed across the environment. The total number of range images was 7 and the total number of points was 781. The average error at each point containing measured and aligned errors was 2.2. As shown in Figure 11 and 12, the results show our proposed method could be applied to actual data containing some noises.



(a) Before alignment (b) Before alignment

Figure 11. Experimental results.



(a) Before alignment (b) Before alignment

Figure 12. Range images by connecting nearest points.

7. Conclusion

In this paper, we describe a new alignment method for reconstruction of a high-resolution 3D shape from low-resolution, partial time-sequential range images and evaluate its performance using some rigid bodies.

References

[1] R. Ginhoux, et al. Beating heart tracking in robotic surgery using 500 Hz visual servoing, model predictive control and an adaptive observer. In *Proc. of 2004 IEEE International Conference on Robotics and Automation*, pages 274–279, 2004.

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

[3] K. Nishino, et al. Robust simultaneous registration of multiple range images. In *Proc. of the 5th Asian Conference on Computer Vision*, pages 23–25, 2005.

[4] S. Rusinkiewicz, et al. Real-time 3D model acquisition. *ACM Transactions on Graphics*, 21(3):438–446, 2002.

[5] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing*, 25(5):578–596, 2007.

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

[7] J. Ziegler, et al. Tracking of the articulated upper body on multi-view stereo image sequences. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2006.