

# 3D SHAPE MEASUREMENT BY SWEEPING HAND-HELD LIGHT STRIPING RANGE FINDER

Hirofumi Nakai<sup>1</sup>, Daisuke Iwai<sup>1</sup>, Kosuke Sato<sup>1</sup>

<sup>1</sup>Graduate School of Engineering Science, Osaka University, Osaka, Japan  
(Tel : +81-6-6850-6372; E-mail: nakai@sens.sys.es.osaka-u.ac.jp, {daisuke.iwai,sato}@sys.es.osaka-u.ac.jp)

**KEY WORDS:** Range Finder, 3D Modeling, Handy Scanner, Slit light projection, ICP Algorithm

## ABSTRACT:

This paper presents a novel hand-held 3D scanning system using a fixed camera and a 1D time-of-flight (TOF) range scanner. A user of the system sweeps the target object with the slit light projected from the scanner. The system obtains sets of the measured range data and captured image of the projected slit light. Our objective is to estimate the rigid-body transform of the scanner in a world coordinate system at each capture frame. To achieve this, we select pairs of corresponding points between the captured slit light and the measured range data, and estimate the transform that minimizes a sum of point-to-point distances of the pairs with an ICP algorithm. Once the transform is estimated, all the range data are transformed in the world coordinate system, and merged into an entire surface shape of the target object. We built a prototype system and conducted proof-of-concept experiments to verify the proposed method.

## 1 INTRODUCTION

Laser range finders are widely used in digital archiving of shape of historical heritages. In such eHeritage research projects, large and heavy equipments are usually applied [Levoy et al. 2000][Ikeuchi et al. 2004]. As an extreme example, Levoy et al. built a motorized gantry on which a laser triangulation scanner was attached to acquire 3D shapes of Michelangelo's statues [Levoy et al. 2000]. The gantry weighed 1,800 pounds and stood 7.5 meters tall, and thus needed large and firm places. The measurements were done on marble floors supported below by massive masonry vaults.

Such large and heavy equipment is not suitable for measuring of relics that exist in narrow spaces (e.g., caves) or on irregular terrains (e.g., outside archaeological sites). In these cases, hand-held shape measurement tools are more useful and convenient. However, even common laser range finders are too heavy for users to hold and carry around in the sites for hours: for example, both VIVID 9i<sup>1</sup> and LMS-Z390i<sup>2</sup> weigh 15 kg.

Recently, several researchers have focused on developing hand-held range scanners for digital archiving of historical heritages at sites in which common scanners cannot easily be installed. In their systems, a user holds a laser slit projector, and sweeps an object with the projected slit light. Once a rigid-body transform, or 6 degree-of-freedom (DOF) of rotation and translation, of the projector is estimated, object's surface shape under the slit light is measured through triangulation. Basically the systems consist of a fixed video camera and a laser slit projector. The camera is used to extract a position of a slit light on the object as well as to estimate the projector's rigid-body transform relative to the camera.

Takatsuka et al. proposed a hand-held 3D scanning system in which LED markers are attached on the spot light projector [Takatsuka et al. 1999]. Positions of LED markers in a captured image are used to calculate 6DOF rotation and translation of the projector at each capture frame. Furukawa et al. proposed a method in which any additional visual markers such as LEDs are not needed [Furukawa et al. 2003]. A relative rigid-body transform of the slit projector is estimated only by the captured slit lights that are projected onto the object.

While various hand-held scanning methods based on triangulation have been reported, no work has been published on one based on time-of-flight (TOF) range finder. Our aim is to propose a 3D shape measurement system that consists of a fixed video camera and a hand-held 1D TOF range finder which also projects a slit light. The proposed system does not need any visual markers for estimation of the range finder's relative rigid-body transform. Because a TOF range scanner itself measures shape of the object under the slit light, estimation of rigid-body transform becomes simpler than a previous work proposed by Furukawa et al. [Furukawa et al. 2003]. First, we extract a projected slit light in the captured image. Second, we select pairs of corresponding points between the captured slit light (2D) and the measured range data (3D), and find a relative rigid-body transform of the range scanner that minimizes an error metric, such as the sum of point-to-point distances. We applied an iterative closest point (ICP) algorithm to refine the transform. Finally, we merge all the range data based on the estimated rigid-body transforms, and obtain entire surface shape of the target object.

The remainder of the paper is organized as follows. The subsequent section briefly describes related studies. Section 3 describes the detailed principles of the proposed method that estimates a rigid-body transform of the range finder relative to the fixed camera. Section 4 presents proof-of-experiments of shape measurements to evaluate the proposed method. In section 5, we briefly discuss and limitations of the proposed method. Section 6 concludes the paper and directions for future work.

## 2 RELATED WORKS

Hand-held shape measurement systems have been developed by several researchers. As mentioned in Sec. 1, Furukawa et al. a system which consists of a fixed camera and a hand-held laser slit project on which multiple LEDs are attached [Furukawa et al. 2003]. Takatsuka et al. proposed almost the same system, but they used a laser spot projector instead of slit projector [Takatsuka et al. 1999]. In these methods, the rigid-body transforms of the projectors are computed by analyzing positions of the captured LED markers. They cannot be used when the LEDs are not captured by the camera. Thus, users have to take into account the view volume of the camera while moving the projector. To solve this problem, Davis et al. and Kawasaki et al. proposed methods

<sup>1</sup>Konica Minolta: <http://www.konicaminolta.com/>

<sup>2</sup>RIEGL: <http://www.riegl.com/>

which do not need any additional visual markers or 6DOF sensors [Davis et al. 2001][Kawasaki et al. 2006].

There are other approaches to achieve hand-held shape measurement. One of the most popular method is structure-from-motion in which a user holds and carries around a camera, and captures multiple images of the target object from different viewpoints. Surface shape of the object is reconstructed by analyzing the image sequence [Pollefeys et al. 1999]. Bouguet et al. and Fisher et al. proposed shape measurement using shadow [Bouguet et al. 1998][Fisher et al. 1999]. In their methods, a user shakes a stick which casts a shadow on the target object and a video camera captures the scene. Shape of the object is reconstructed by the image sequence. Rusinkiewicz et al. and Hebert et al. proposed methods in which users hold the target object in stead of the raser scanner, and move and rotate it in front of the raser scanner to obtain entire surface shape of the object [Rusinkiewicz et al. 2002][Hebert 2001].

### 3 3D MODELING PROSESS

#### 3.1 System overview

Figure 1 shows our system configuration. The system consists of a fixed video camera and a hand-held range scanner. The scanner emits 1 slit of light on the target object. Distances from the scanner to the object surface on which the light is projected are measured based on TOF method at a video rate. A user holds the range scanner and sweeps object with the projected slit light to measure the entire surface shape of the object.

The fixed video camera observes the projected light. The origin of the world coordinate system is assigned to the center of a lens. We estimate a rigid-body transform of the laser scanner in the world coordinate system at each video frame by comparing obtained range data with a corresponding slit image. Note that the captured image of the projected slit light is referred to as ‘slit image’.

Figure 2 shows the process flow of the proposed system. First, a set of a slit image and range data is acquired at each video frame. Second, for each set, the proposed system finds pairs of corresponding points between the captured slit light (2D) and the measured range data (3D), and then estimates a rigid-body transform of the scanner through ICP algorithm. This process is referred to as ‘matching process’ and consists of two steps: global matching and local matching. The details of the method are described in the following sections. Once the rigid-body transform is estimated, each range data are geometrically transformed into the world coordinate system. Finally, all the transformed range data are integrated and consequently entire surface shape of the target object is obtained.

#### 3.2 Principle of rigid-body transform estimation of the laser scanner

In principle, a rigid-body transform of the scanner is estimated based on perspective projection transformation.

The relationship between the camera coordinate system  $\mathbf{M}_c = [x_c y_c z_c]^T$  and the range scanner coordinate system  $\mathbf{M}_r = [x_r y_r z_r]^T$  are represented as:

$$\mathbf{M}_c = \mathbf{R}\mathbf{M}_r + \mathbf{t}, \quad (1)$$

where  $\mathbf{R}$  is a  $3 \times 3$  rotation matrix and  $\mathbf{t}$  is a 3DOF translation vector. When a projected light is captured by the camera, the relationship between the 3D range data and the 2D position of the

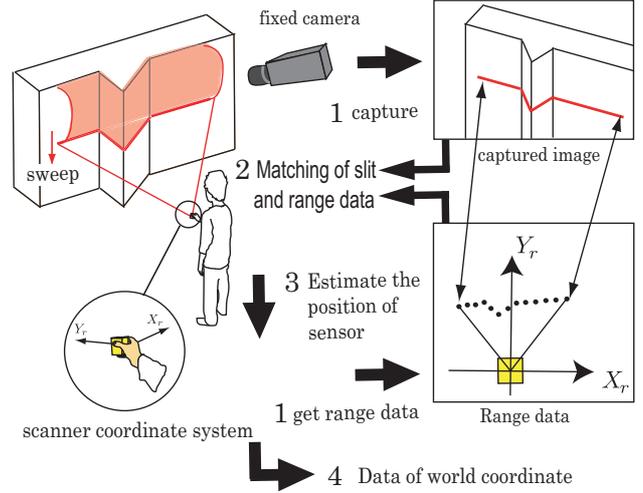


Figure 1: System overview: 1. a set of a slit image and range data is acquired at each video frame. 2. For each set, the proposed system finds pairs of corresponding points between the captured slit light (2D) and the measured range data (3D). 3. Estimates a rigid-body transform of the scanner. 4. Transformed range data in the camera coordinate system

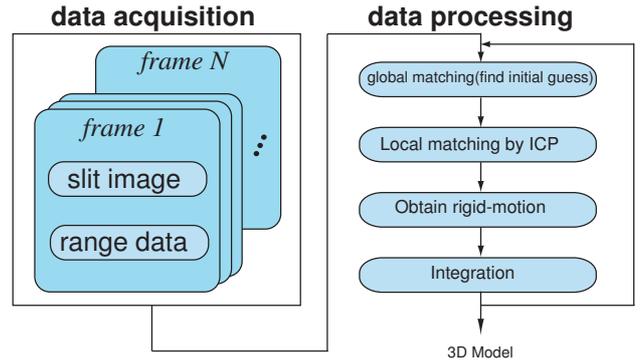


Figure 2: Process flow

projected light in the captured image is represented as the following equation based on the perspective projection transformation:

$$\lambda \tilde{\mathbf{m}}_c = \mathbf{P}\tilde{\mathbf{M}}_r = \mathbf{A} \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{T}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \tilde{\mathbf{M}}_r = \mathbf{A}\mathbf{D}\tilde{\mathbf{M}}_r \quad (2)$$

where  $\tilde{\mathbf{m}}_c = [u \ v \ 1]^T$  is the position vector in the camera screen coordinate system, and  $\tilde{\mathbf{M}}_r = [x_r \ y_r \ z_r \ 1]^T$  is one in the range scanner coordinate system.  $\mathbf{P}$ ,  $\mathbf{A}$  and the  $\mathbf{D}$  are  $3 \times 4$  perspective matrix, the  $3 \times 4$  camera’s intrinsic matrix, and its extrinsic matrix, respectively.  $\mathbf{D}$  represents arbitrary 6DOF rotation and translation.

Since we apply a 1D range scanner, the Z axis of the range scanner coordinate system can be ignored. Thus, the equation (2) gives:

$$\lambda \tilde{\mathbf{m}}_c = \mathbf{P}' [X_r \ Y_r \ 1]^T, \quad (3)$$

where  $\mathbf{P}' \in \mathbf{R}_{3 \times 3}$  is a matrix obtained by reducing the third column of  $\mathbf{P}$ , and  $\tilde{\mathbf{M}}'_r = [x_r \ y_r \ 1]^T$ . Suppose that there are n correspondences,  $\mathbf{P}'$  can be calculated as follows:

$$\mathbf{B}\mathbf{p} = \mathbf{q}, \quad (4)$$

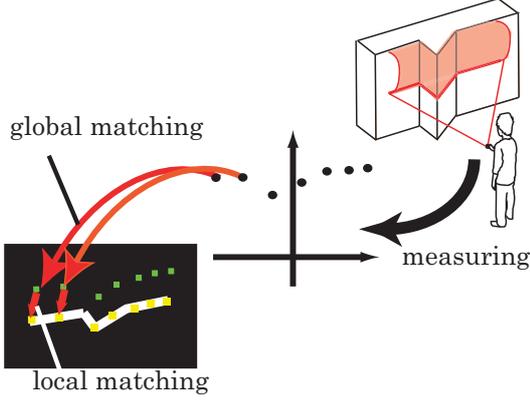


Figure 3: Matching of range data and slit light image

$$\mathbf{B} = \begin{bmatrix} x_{r1} & y_{r1} & 1 & 0 & 0 & 0 & -ux_{r1} & -uy_{r1} \\ 0 & 0 & 0 & x_{r1} & y_{r1} & 1 & -vx_{r1} & -vy_{r1} \\ \vdots & \vdots \\ x_{rn} & y_{rn} & 1 & 0 & 0 & 0 & -ux_{rn} & -uy_{rn} \\ 0 & 0 & 0 & x_{rn} & y_{rn} & 1 & -vx_{rn} & -vy_{rn} \end{bmatrix}, \quad (5)$$

$$\mathbf{P} = [p_{11} \ p_{12} \ p_{14} \ p_{21} \ p_{22} \ p_{24} \ p_{31} \ p_{32} \ p_{34}]^T, \quad (6)$$

where  $p_{ij}$  is each element of the matrix  $\mathbf{P}$ .  $\mathbf{P}$  can be estimated through a least square method by taking more than four corresponding points between  $\mathbf{m}_c$  and  $\mathbf{M}_r$ . Once  $\mathbf{P}$  is estimated,  $\mathbf{D}$  can be calculated if the camera's intrinsic matrix  $\mathbf{A}$  is known. We obtained the intrinsic matrix using Zhang's method in advance [Zhang 2001].

### 3.3 Matching of range data and slit image

To estimate the extrinsic matrix  $\mathbf{D}$ , pairs of corresponding points between the captured slit light (2D) and the range data (3D) have to be found. The matching process consists of two stages: global and local matching (Fig. 3). In the global matching, we estimate the initial matching by using end points of the slit light in both data. In the local matching, the range data is firstly projected on the slit image according to the rigid-body transform estimated in the global matching. Then, for each point of the projected range data, we select a point on the slit light in the captured image so that the selected point is closest to it. We applied an ICP algorithm to the local matching process.

**3.3.1 Global matching** In the global matching, a simple 2D affine transformation is applied because both the range scanner and camera screen coordinate systems are 2D. At first, the end points of the slit light are extracted in both the data. We define 2D base vectors  $\mathbf{e}_c$  (camera screen) and  $\mathbf{e}_r$  (range scanner) which connect the end points of the slit light in both the data.  $\mathbf{n}_c$  and  $\mathbf{n}_r$  are the normal vector of  $\mathbf{e}_c$  and  $\mathbf{e}_r$ . Note that these vectors are normalized. Then orthogonal matrices  $\mathbf{E}_c = [\mathbf{e}_c \ \mathbf{n}_c]^T$  and  $\mathbf{E}_r = [\mathbf{e}_r \ \mathbf{n}_r]^T$  are defined with the vectors. We calculate the gravity points of the slit light as  $\mathbf{m}_{gc} = [x_{gc} \ y_{gc}]^T$  and  $\mathbf{m}_{gr} = [x_{gr} \ y_{gr}]^T$ . Then, equation. (8) can transform the two coordinate systems.

$$\mathbf{E}_c^T(\mathbf{m}_c - \mathbf{m}_{gc}) = \mathbf{S}\mathbf{F}_y\mathbf{E}_r(\mathbf{m}_r - \mathbf{m}_{gr}), \quad (7)$$

$$\mathbf{F}_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$$

where  $\mathbf{m}_c = [x_c \ y_c]^T$  and  $\mathbf{m}_r = [x_r \ y_r]^T$  are new points on the base vector  $\mathbf{e}_c$  and  $\mathbf{e}_r$ , respectively.  $\mathbf{S}$  is a scale coefficient and  $\mathbf{F}_y$  is a matrix which flips the y-axis of the range scanner coordinate system.  $\mathbf{m}_c$  (or  $\mathbf{m}_r$ ) is selected so that a distance between  $\mathbf{m}_c$  (or  $\mathbf{m}_r$ ) and the base vector  $\mathbf{e}_c$  (or  $\mathbf{e}_r$ ). Once estimating the above parameters, we can calculate the initial matching through 2D affine transformation as follows:

$$\mathbf{m}_c = \mathbf{E}_c\mathbf{S}\mathbf{F}_y\mathbf{E}_r(\mathbf{m}_r - \mathbf{m}_{gr}) + \mathbf{m}_{gc}, \quad (8)$$

$$= \mathbf{R}\mathbf{x}_r + \mathbf{t}, \quad (9)$$

$$\mathbf{R} = \mathbf{E}_c\mathbf{S}\mathbf{F}_y\mathbf{E}_r, \quad (10)$$

$$\mathbf{t} = \mathbf{E}_c\mathbf{S}\mathbf{F}_y\mathbf{E}_r\mathbf{m}_{gr} + \mathbf{m}_{gc}. \quad (11)$$

We go through this process only once in the first frame. For the rest, the extrinsic matrix  $\mathbf{D}$  that is estimated in the previous frame is used as the initial guess.

**3.3.2 Local matching** We applied an ICP algorithm to accurately select pairs of corresponding points between the captured slit light and the measured range data.

The ICP algorithm starts with the initial guess, and iteratively refines the transform by repeatedly generating pairs of corresponding points and finding the transformation that minimize an error metric. The ICP generally consists of the following six steps:

1. **Selection** of points in one or both data set
2. **Matching** these points to samples in the other data set
3. **Weighting** the corresponding pairs
4. **Rejecting** certain pairs
5. Assigning an **error metric** to the current point pairs
6. Finding a rigid-body transform which **minimizes** the error metric

In our case, we have applied the followings:

1. All points in the range data are selected.
2. The selected points are projected on the camera screen coordinate system. For each projected point, a point on the slit light in the captured image is matched if the distance between these points is the closest and the angle between their normal vectors is within 45 degree.
3. A uniform value is weighted to all the points.
4. End points of the slit light are rejected.
5. A sum of point-to-point distances is applied as the error metric. Suppose that the corresponding points of the  $i$ -th pair are represented as  $\mathbf{M}_i$  (range data) and  $\mathbf{m}_i$  (slit image), the error is represented as the following equation:

$$E = \Sigma \left\| \frac{1}{\lambda} \mathbf{A}(\mathbf{R}\mathbf{M}_i - \mathbf{t}) - \mathbf{m}_i \right\|. \quad (12)$$

6. A rigid-body transform which minimizes the error metric is found through the Levenburg-Marquardt algorithm.

Consequently, the optimal pairs of corresponding points between the capture slit light and the range data are obtained.

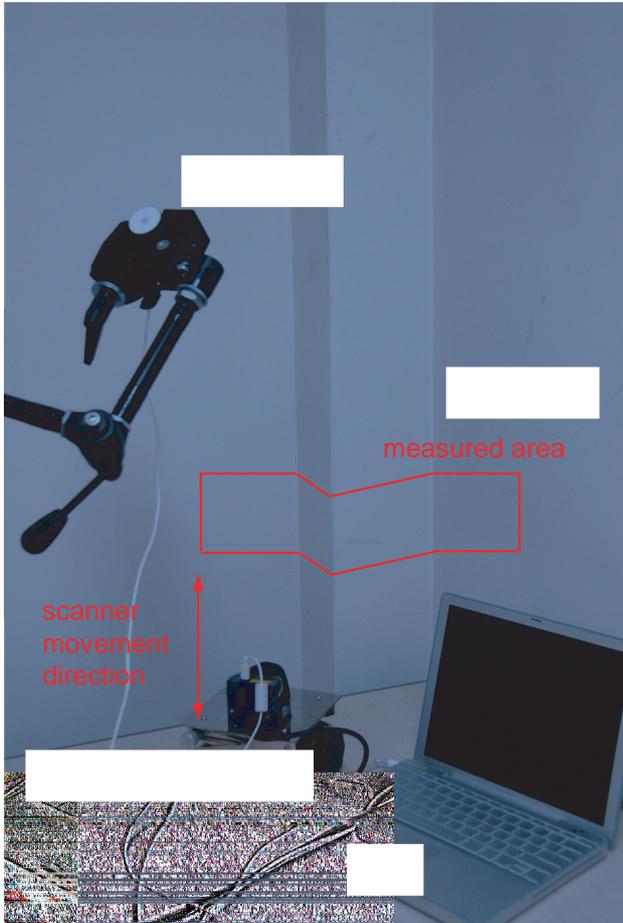


Figure 4: System setup of our prototype

## 4 EXPERIMENT

### 4.1 System setup

To verify our method, we built the prototype system and demonstrate 3D model acquisition. Our experimental setup is shown in Fig. 4. The system consists of a video camera and a 1D scanning range scanner. The range scanner emits a slit of near-infrared (NIR) laser and measures distance between the scanner and the target object based on TOF method (Fig. 5 (a)). Specifications of the scanner is shown in Table 1. Distances of 682 points in the angle of 240 degrees are measured at 28 ms/scan. We use 129 points out of the 682 points which correspond to the angle of 60 degrees. The video camera is sensitive to wavelengths of NIR and captures the projected scene at 30 fps. Fig. 5 (b) shows an example of the NIR image of the laser slit that is projected onto a corner of our laboratory. We applied thresholding and thinning processes to the slit image and extracted the slit line (Fig. 5 (c)). the range data is shown in Fig. 5(d). The range data had error variance of  $\pm 20$ mm.

### 4.2 Results

We measured a surface shape of corner of a building. We moved the scanner vertically (Fig. 4) and obtained 13 range data (totally 1677 points) and corresponding scan images. For each data, we estimated the 6DOF rotation and translation of the scanner. The range data are geometrically transformed and merged in the world coordinate system. Figure 6 (a), (b) show the merged surface shape of the corner that are rendered from different view-

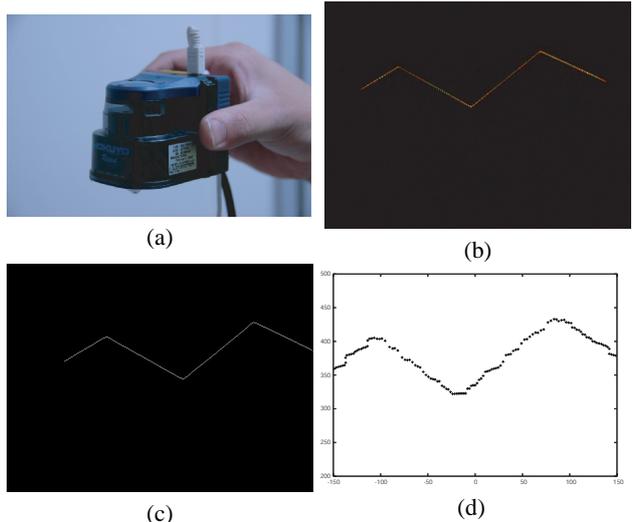


Figure 5: (a) Range scanner, (b) Detected laser stripe, (c) Detected slit stripe performed by thresholding and thinning, (d) Sample of range data

Table 1: The specifications of the laser scanner

practical of measuring	Time-of-flight method
measuring velocity	28 ms/scan
angle of scanning	240 degree
measuring accuracy	$\pm 10$ mm
resolution	682 resolution/240 degree
laser class	1
manufacturer	HOKUYO AUTOMATIC CO.
dimension	60 mm x 60 mm x 75 mm
weight	185 g

points. We confirmed that there is no significant misalignment of the range data.

Figure 7 shows the trajectory of the range scanner in the world coordinate system. The result shows that the scanner moves in one direction, thus we confirmed that the rigid-body transform (particularly translation) of the scanner is correctly estimated.

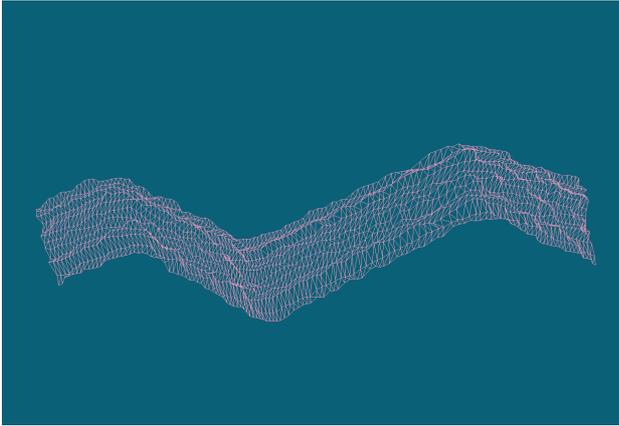
Figure 8 shows the sequence of the iterations of the local matching. The white points represent the range data that are projected on the slit image, and the white line represents the extracted slit light in the slit image. It can be seen that these two data are getting closer each other through the iteration process. Figure 9 shows the error metric in each iteration step. The error (sum of the point-to-point distances) and number of iterations are assigned to the vertical and horizontal axes, respectively. There are 9 iterations until convergence. The final error becomes only a quarter of the initial one.

Figure 11 (a), (b) show results of Vietnam hat (Fig. 10). We obtained 60 range data (totally 7740 points) and merged in the world coordinate system.

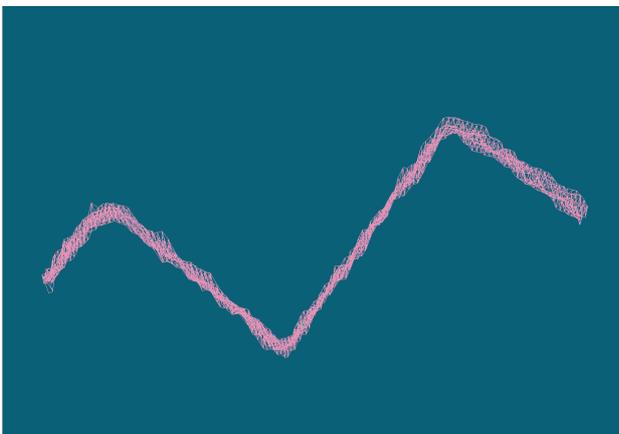
## 5 DISSCUSION

### 5.1 Limitation of rigid-body transform estimation

When the slit light in the captured image becomes straight (e.g., the target object is a planar surface), the estimation of the rigid-body transform fails. This is because that the matrix  $\mathbf{B}$  of equation (4) becomes rank deficient. To solve this problem, additional



(a)



(b)

Figure 6: Modeling results rendered in wireframe (a) bird's view, (b) top view

constraints are needed. We consider that intersections among the slit lights will give significant constraints. Improvement of the rigid-body transform estimation based on the slit light intersections is to be done in our future work.

## 5.2 Limitation of slit extraction

A vital assumption of our method is that the slit light is correctly extracted in the captured image. If there are some noises or discontinuities of the slit light in the image, the matching process fails. We apply a background subtraction method to extract the slit light in the captured image. Thus, the preciseness of the measured shape data depends on the environmental light condition. While the proposed algorithm works well in a dark place such as an underground passage in a heritage, it is quite hard to acquire a precise 3D model in outdoor because noises caused by a rapid change of the environmental light in the measurement cannot be ignored. An analytical evaluation of the relationship between the noise level and the quality of the final surface shape, and a development of a more robust algorithm would be one of the future works.

## 6 CONCLUSIONS

We proposed a hand-held shape measurement system using a 1D TOF range scanner and a fixed video camera. In the system, additional sensors are not required to measure rigid-body transform of the scanner. Instead, the proposed system estimates the transform

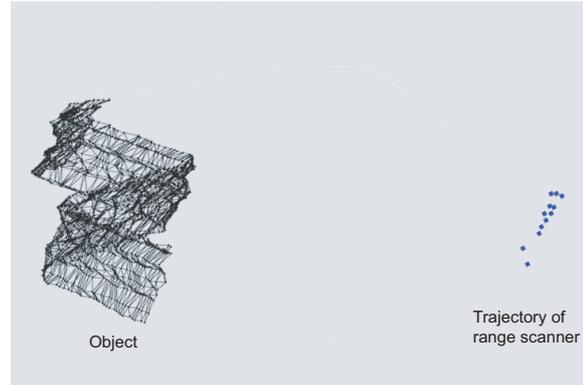


Figure 7: Trajectory of the range scanner. Blue points show the position of the range scanner respectively.

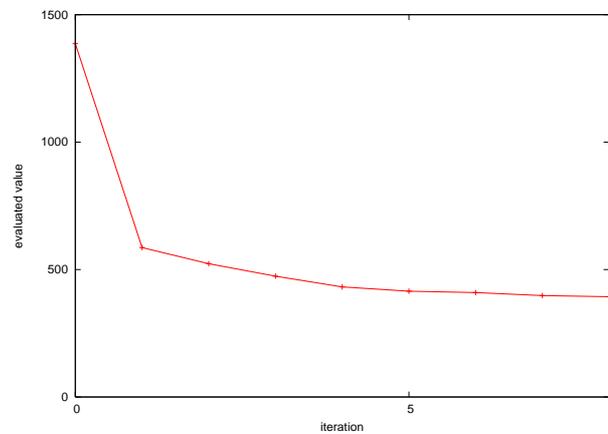


Figure 9: Convergence of value E

by matching the captured slit light and the obtained range data through ICP algorithm. We built the first prototype and conducted the initial proof-of-concept experiments. Because of a limitation on the current 3D recovering algorithm, the proposed system is able to measure only objects with continuous surfaces. However, the system may support for archaeological field works to quickly visualize the local detail shape of structures in rough condition.

## REFERENCES

- M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk: he digital michelangelo:3d scanning of large statues, SIGGRAPH 2000, Proceedings, pages. 131-144, 2000.
- K. Ikeuchi, K. Hasegawa, A. Nakazawa, J. Takamatsu, T. Oishi and T. Masuda: Bayon Digital Archival Project, TICVSM, Proceedings, pages. 334-343, 2004.
- H. Kawasaki, R. Furukawa: Interactive Shape Acquisition using Marker Attached Laser Projector, 3DIM, Proceedings, pages. 491-498, 2003.
- H. Kawasaki, R. Furukawa: Self-calibration of Multiple Laser Planes for 3D Scene Reconstruction, Proceedings, 3DPVT, pages. 200-207, 2006.
- M.Takatsuka, G.A.W.West, S.Vetterling, T.M.Caelli,: Low-cost interactive active monocular range finder, CVPR, Proceedings, pages. 444-449, 1999.

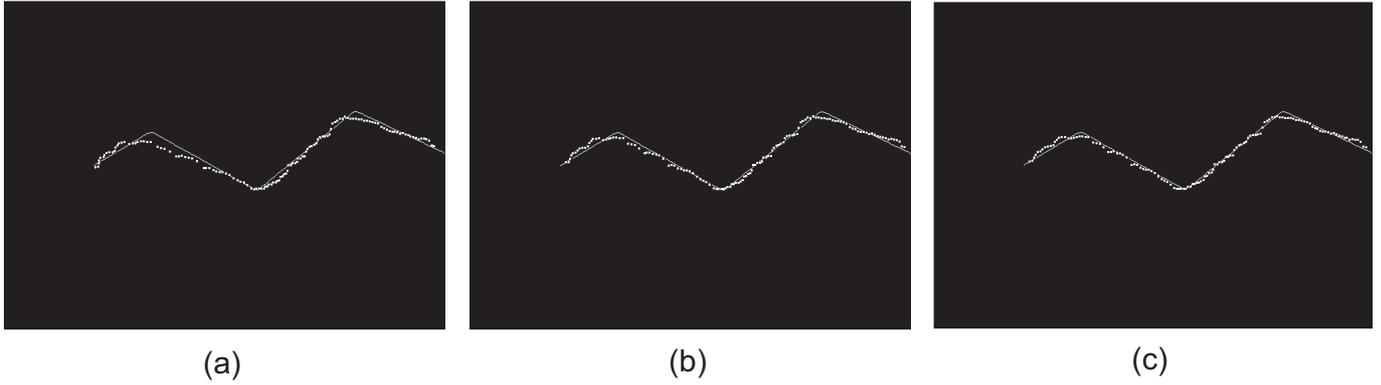


Figure 8: Convergence of local matching process (a) initial condition (result of global matching), (b) 4 iterations, (c) 9 iterations

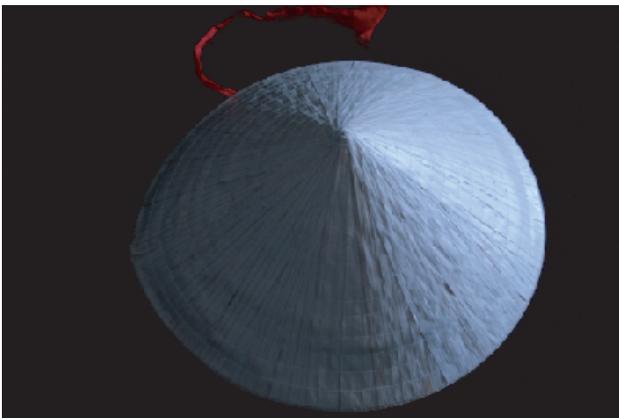


Figure 10: Target object (Vietnam hat)

J. Y. Bouguet and P. Perona: 3D: photography on your desk, CV, Proceedings, pages. 129-149, 1998.

S. Rusinkiewicz, O. Hall-Holt, M. Levoy: Real-Time 3D Model Acquisition, SIGGRAPH, Proceedings, pages. 438-446, 2002.

P. Hebert: A self-Referenced Hand-Held Range Sensor, 3DIM, Proceedings, pages. 5-12, 2001.

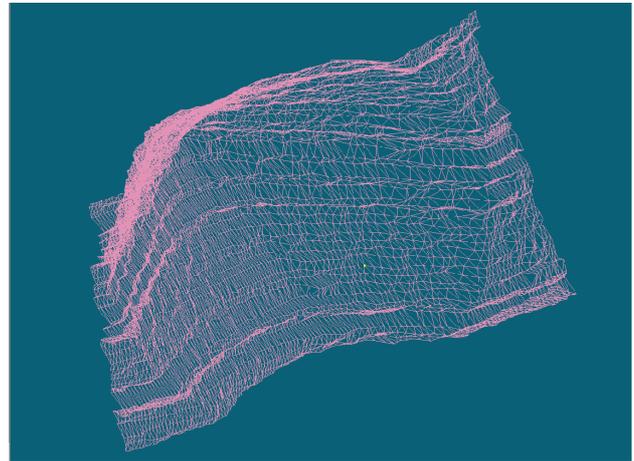
R. Fisher, A. Ashbrook, C. Robertson, N. Werghi: A Low-Cost Range finder using a Visually Located, Structured Light Source, 3DIM, Proceedings, pages. 24-33, 1999.

M. Pollefeys, R. Koch, M. Vergaumen, L. Van Gool: Hand-Held acquisition of 3D Models with video camera, 3DIM, Proceedings, pages. 14-23, 1999.

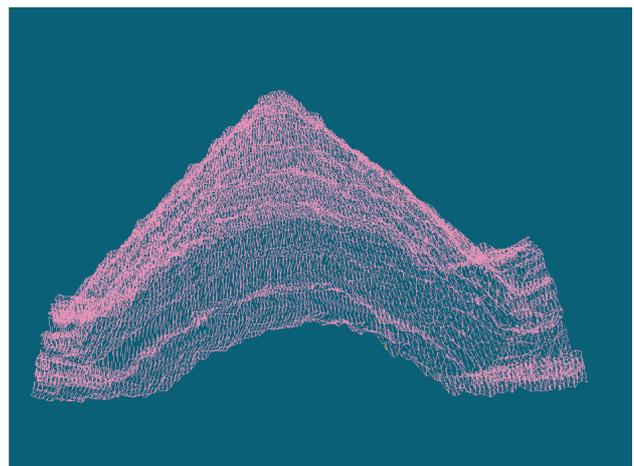
J. Davis, X. Chen: A laser range scanner designed for minimum calibration complexity, 3DIM, Proceedings, pages. 91-98, 2001.

P. J. Besl and N. D. McKay: A method for registration of 3-d shapes, PAMI, Transaction, Vol. 14, No. 2, pages. 239-256, 1992.

Z. Zhang: Flexible Camera Calibration By Viewing a Plane From Unknown Orientations, ICCV, Technical Report, pages. 666-673 1998.



(a)



(b)

Figure 11: Modeling results of Vietnam hat in Fig. 10