

Surface Extraction from Incomplete Point Data for Digitizing Factories

Hiroshi Masuda*
The University of Tokyo

Satoshi Fujii†
The University of Tokyo

Abstract

Recently, mid-range and long-range laser scanners for measuring large-scale facilities have made remarkable progress. In this paper, we discuss reverse engineering of large engineering facilities using those scanners. Our goal is to estimate shape parameters of components in engineering facilities. Although the state-of-the-art laser scanners can produce hundreds of millions point data in several minutes, those data tend to include large noises and quite a lot of outliers. In addition, the locations of scanners are restricted in most engineering facilities, and therefore components in facilities are measured only from one side and sometimes are partially occluded by other ones. In this paper, we first investigate robust surface estimators for smoothing point-clouds and compare the efficiency and quality of four smoothing methods. Then, we will show how to extract shape parameters as precisely as possible by considering industrial standards and assembly constraints.

Keywords: digital factory, point-cloud, smoothing, robust estimate, surface fitting

1 Introduction

The rapid progress of laser scanners has accelerated the research on reverse engineering. However, most research activities in the CAD community concentrate on relatively small mechanical parts that are measured by triangulation-based scanners. Triangulation-based scanners can precisely measure mechanical parts, but they typically cover only within a range of a few meters. Therefore, triangulation-based scanners are not suitable for measuring large objects such as factories, power plants and logistics centers.

In recent years, 3D shape acquisition of large engineering facilities, such as industrial plants and power plants, has been receiving increasing attention in surveying communities. One of useful applications is 3D simulation of complicated maintenance and repair tasks. It is widely recognized that model-based planning based on 3D CAD reduces the rework of maintenance tasks to a large extent. Currently, one of the biggest obstacles that prevent model-based simulation is that old facilities lack not only 3D CAD models but also reliable drawings, because many engineering facilities were built one or more decades ago and have been repeatedly renovated in their long lifecycles. In such cases, it is very useful to generate 3D as-built models by measuring real facilities.

While triangulation-based scanners are popular in the CAD community, mid-range and long-range laser scanners have made remarkable progress in the field of surveying. Mid/long-range scanners have been mainly used for surveying large-scale facilities, geography and infrastructures. We intend to use mid/long-range scanners for creating 3D models for digital factories.

Two types of scanners, the time-of-flight scanners and the phase-based scanners, are typically used for surveying.

The time-of-flight scanner measures the round-trip travel time of the laser pulses. This type of scanner is called a long-range scanner, which can measure in the range of a few hundred meters. However, it takes a lot of time to measure many points that cover large facilities, because the time-of-flight scanner must wait a round-trip laser pulse for each point.

The phase-based scanner is called a mid-range scanner. The state-of-the-art phase-based laser scanners can measure objects up to 79 meter [1] or 120 meter [2]. This type of scanners radiates continuous modulated laser pulses and calculates distances using the phase difference between the emitted and received signals. One great advantage of this type of scanner is that it can measure two hundred million points in several minutes. These measuring range and speed allow to measure manufacturing factories in non-operating hours, and therefore we believe that the phase-based scanners are sufficient for measuring most engineering facilities.

For modeling components in engineering facilities, it is important to extract surface equations of components and to determine their shape parameters as precisely as possible. This is because most components in engineering facilities are industry standard components, which consist mostly of simple surfaces, such as planes, cylinders, cones, spheres, and tori. For this purpose, we have to solve the following problems.

- (1) The detection of surface regions using very noisy point data.
- (2) The calculation of surface equations using incompletely scanned point data.

Since mid-range and long-range scanners produce very noisy point data compared to triangulation-based scanners, it is not easy to precisely extract shape parameters. Especially, 3D point data acquired by the phase-based laser scanner include large noises along the laser pulses. Figure 1 shows point data captured from a power plant and their mesh model. Since a certain rate of distance errors are greater than 6 mm, the surface of a mesh model often becomes “prickly,” as shown in Figure 1(c).

Region-growing [3, 4] is a typical surface extraction method for detecting planes, cylinders, cones, spheres and tori in point-clouds. This method starts from a small seed region and extends the boundary of the region by checking if neighbor points satisfy the surface equation. Since noisy data prevent to grow regions, it is highly required to smooth point data. Other segmentation techniques [5, 6] have also difficulties in cases of noisy data, because they use local differential properties, which can be calculated on smooth surfaces.

In addition, the locations of a scanner are restricted in engineering facilities. As a matter of course, it is not allowed to disassemble factories for surveying components. Therefore most components are measured only from one or two sides and some components are partially occluded by other ones. Special care must be taken for stably calculating shape parameters.

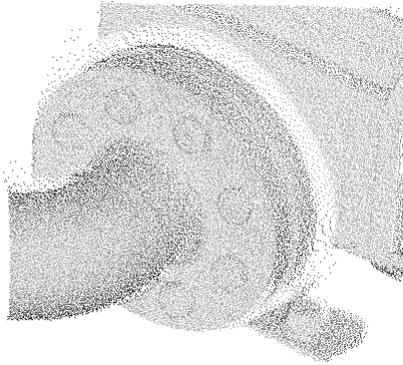
In this paper, we introduce robust estimators for smoothing very noisy point-clouds and compare the efficiency and quality of the

*e-mail: masuda@sys.t.u-tokyo.ac.jp

†e-mail: satosi.fujii@gmail.com

methods. Then, we will discuss how to extract shape parameters as precisely as possible.

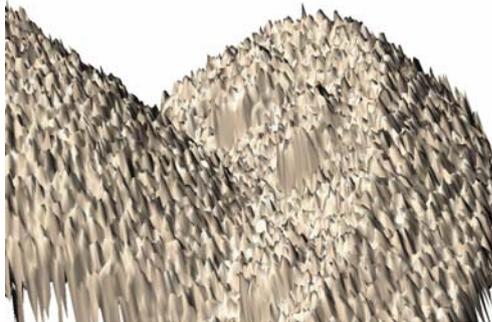
In the following section, we will explain smoothing operators based robust estimate. Section 3 presents how to detect shape parameters from incomplete point data. Section 4 states conclusions.



(a) Point data



(b) Noisy mesh model



(c) Close-up of the noisy mesh model

Figure. 1: Noisy mesh model constructed from point data captured by a phase-based scanner.

2 Smoothing of Noisy Point Data

2.1 Smoothing of Scanned Point Data

The moving-least-squares (MLS) method is a popular tool for generating smooth surfaces [7–9]. The MLS projection calculates a locally smooth surface around each point and projects the point onto the smooth surface.

However, MLS projection often fails to preserve geometric features, when outliers and high levels of noise are involved in point data. Figure 2 shows a result of MLS projection. As shown in this figure, smooth planes become bumpy and sharp edges and corners are not preserved. This is because the mid/long range

scanner produces much more outliers and much larger noises than the triangulation-based scanner. Point data in this paper were captured by Z+F Imager 530 [1], which is a mid-range phase-based scanner.

For smoothing noisy point data, we introduce smoothing methods based on robust estimate. Roughly speaking, robust estimate reduces impacts of outliers by assigning small weights to them. In our previous paper [10], we introduced robust estimate based on Lorentzian distribution. In this paper, we implement other robust estimators and compare their efficiency and quality.

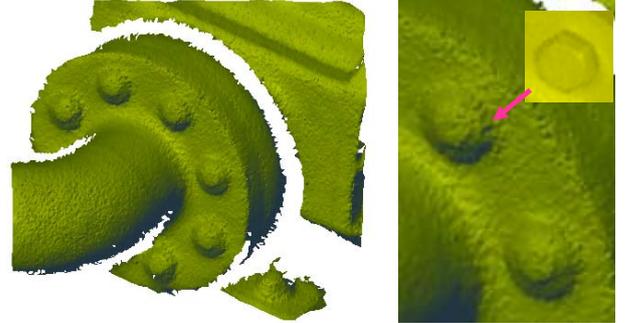
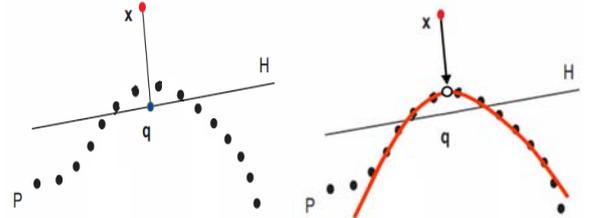


Figure.2: Smoothing by Conventional MLS projection



(a) Projection onto plane H (b) Projection onto a quadratic surface
Figure. 3: MLS projection

2.2 Conventional MLS Projection

We will briefly explain the conventional MLS projection [7, 8] using Figure 3. This method smoothes point x as follows:

- (1) Neighbor points of x are detected.
- (2) Approximate plane H is calculated (Figure 3(a)).
- (3) Reference point q is calculated by projecting x on H .
- (4) Neighbor points are projected on H and a height field is generated.
- (5) A quadratic surface is fitted to the height field (Figure 3(b)).
- (6) The original x is projected on the surface.

Let $S(\mathbf{x}|\mathbf{a})=0$ be the equation of a quadric polynomial surface, which is used to locally approximate the neighborhood of point x . $\mathbf{a}=\{a_k\}$ ($1 \leq k \leq M$) denotes parameters of the surface. M is the number of parameters.

Then, the quadratic surface is calculated by:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \left(\sum_{j \in N} \frac{S(\mathbf{p}_j|\mathbf{a})^2}{\sigma^2} \chi(\|\mathbf{p}_j - \mathbf{q}\|) \right), \quad (2.1)$$

where σ is the standard deviation of $S(\mathbf{x}_i|\mathbf{a})$; N is the neighbor points; \mathbf{q} is a reference point. χ is a monotonically decreasing function typically defined by the Gaussian:

$$\chi(d) = e^{-\frac{d^2}{h^2}}, \quad (2.2)$$

where h is a scaling parameter reflecting the spacing between neighboring points.

It is well-known that the least-squares method is sensitive to outliers. MLS projection is effective only when the number of outliers is relatively small. Therefore, we introduce robust surface estimators, which can stably calculate quadratic surfaces even when many outliers are included in point data.

2.3 Robust Estimate

For robustly smoothing noisy point-clouds, we assume probability density functions for the distribution of residuals and then calculate the maximum likelihood surfaces. This approach can be formalized as M-estimate [11-13]. While the conventional MLS projection implicitly assumes the normal distribution, which is sensitive to large noises and outliers, we design a new smoothing operator on the basis of more robust distribution functions.

Let points $\mathbf{x}_i \in \mathbb{R}^3$ ($1 \leq i \leq N$) be sampled from a smooth surface.

$$r_i = \frac{S(\mathbf{x}_i | \mathbf{a})}{\sigma} \quad (1 \leq i \leq N), \quad (2.3)$$

where σ is the standard deviation of the values of $S(\mathbf{x}_i | \mathbf{a})$. Given the probability density function $f(r)$, the maximum likelihood estimate $\hat{\mathbf{a}}$ can be calculated as:

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} \prod_{i=1}^N [f(r_i)] \quad (2.4)$$

or equivalently,

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \sum_{i=1}^N \rho(r_i), \quad (2.5)$$

where $\rho(r_i) = -\log(f(r_i))$.

We introduce a new MLS-like surface estimator based on robust estimate. The smooth surface is then calculated as:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \left(\sum_{j \in N_i} \rho \left(\frac{S_i(\mathbf{p}_j | \mathbf{a})}{\sigma_i} \right) \chi(\|\mathbf{p}_j - \mathbf{c}_i\|) \right). \quad (2.6)$$

When $\rho(r)$ is differentiable, $\hat{\mathbf{a}}$ can be solved by:

$$\frac{\partial}{\partial a_k} \left(\sum_{i=1}^N \rho(r_i) \right) = \frac{1}{\sigma} \sum_{i=1}^N w_i(r_i) \left(\frac{\partial S(\mathbf{x}_i | \mathbf{a})}{\partial a_k} \right) = 0, \quad (2.7)$$

where

$$w_i(r) \equiv \frac{\partial \rho}{\partial r}(r) \quad (2.8)$$

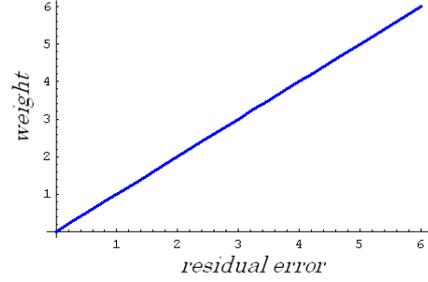
is a weighting function of point \mathbf{x}_i .

Figure 4(a) shows a weight function of normal distribution, which is assumed by the conventional MLS projection. This weight function implies that the larger residuals have larger weights, and therefore, the impacts of outliers are amplified. This is the reason why outliers must be carefully eliminated when MLS projection is applied to point data.

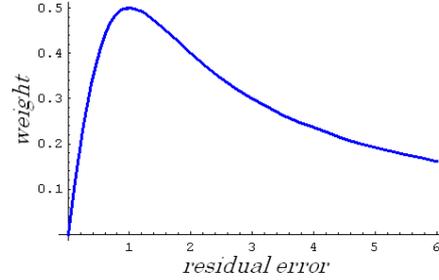
2.4 Lorentzian Surface Estimator

Suppose the Lorentzian distribution $\rho(r) = \log(1+r^2)$ for Equation 2.6. Then, the maximum likelihood surface can be calculated as:

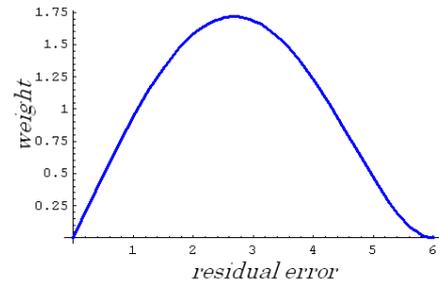
$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \left(\sum_{j \in N_i} \log \left\{ 1 + \frac{S_i(\mathbf{p}_j | \mathbf{a})^2}{\sigma_i^2} \right\} \chi(\|\mathbf{p}_j - \mathbf{c}_i\|) \right). \quad (2.9)$$



(a) Normal distribution (MLS)



(b) Lorentzian distribution



(c) Tukey's bi-weight

Figure 4: Weight functions for surface estimation

The weight function of the Lorentzian distribution is:

$$w_i(r) = \frac{2r}{1+r^2}. \quad (2.10)$$

Figure 4(b) shows the weight function of the Lorentzian distribution. This weight function implies that larger residuals have smaller weights, and outliers will have no practical impact on the estimation of parameters. Therefore, surface estimation based on Equation 2.9 is more robust than conventional MLS projection.

2.5 Surface Estimation by Tukey's Bi-Weight

As shown in Figure 4(a)(b), the robustness of smoothing operators are mainly determined by the design of weight functions. Various robust estimators can be designed so that large residuals have small weights. Tukey designed the following weight function using polynomial curves shown in Figure 4(c):

$$w_i(r) = \begin{cases} r(1-r^2/c^2)^2 & |r| < c \\ 0 & |r| > c \end{cases} \quad (2.11)$$

This function is called the Tukey's bi-weight function, which is not based on any well-known distribution density functions. When points have residuals that exceed c , their weights become 0. The value of c is

typically 6, which means that points in 6σ are only considered for surface estimation.

2.6 Smoothing on a Spherical Height Field

The distance image, which maintains the distance from the origin at each pixel, is popular in computer vision community. In the case of mid-range and long-range scanners, the direction of a laser beam can be represented using an azimuth angle θ and an elevation angle ϕ , as shown in Figure 5(a)(b), and all measured points can be mapped on the $\theta - \phi$ plane and be represented as (r, θ, ϕ) , where r is the distance in the direction (θ, ϕ) (Figure 5(c)). Since each pixel in Figure 5(c) has a distance, the image can be regarded as a distance image.

Distance images can be smoothed by locally fitting quadratic surfaces to distances on the $\theta - \phi$ plane. We apply a robust estimator based on the Tukey's bi-weight to the distance image in Figure 5(c) and investigate the quality and efficiency.

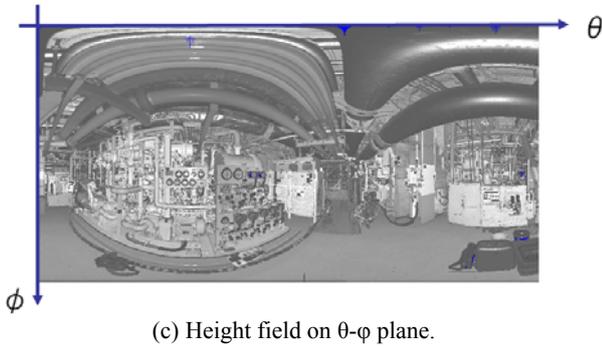
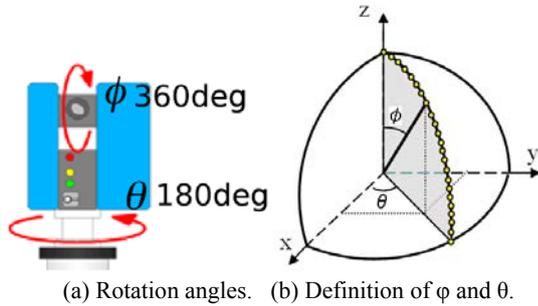


Figure 5: Height field generated by point data captured by a phase-based scanner.

2.7 Experimental Results

Figure 6 shows experimental results of smoothing operators. Mesh models were generated by Delaunay triangulation on the $\theta - \phi$ plane using smoothed point-clouds.

Figure 6(a) is the result of the conventional MLS projection. As shown previously, this method produced a bumpy model.

Figure 6(b) was smoothed by the Lorentzian surface estimator and 6(c) by the Tukey's bi-weight. The both cases preserved features better than the conventional MLS projection.

Figure 6(d) was smoothed as the distance image using Tukey's bi-weight. As shown in this figure, this method could not produce smooth surfaces. In our experiments, the smoothing on the reference planes was better than the one on the $\theta - \phi$ plane.

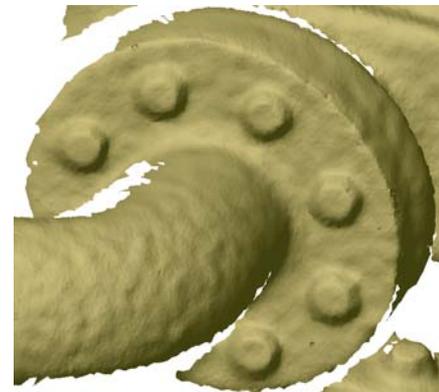
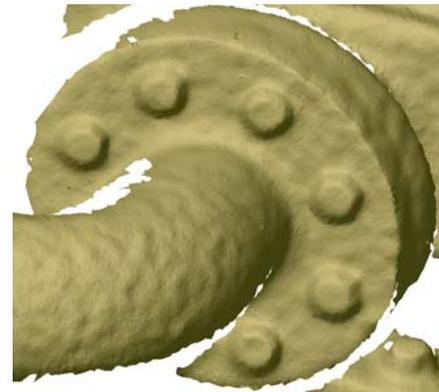
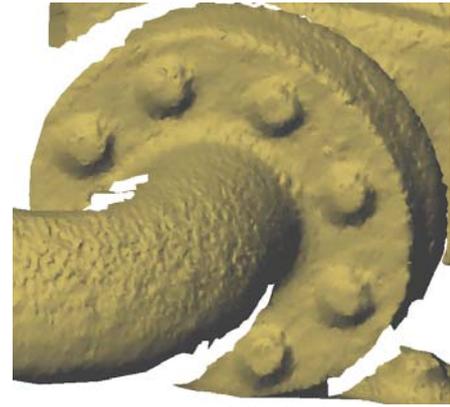


Figure 6: Results of smoothing operators

Table 1 shows how many points could be processed by smoothing operators. The result shows robust estimate based on the Lorentzian distribution was the most time-consuming, and the conventional MLS projection was the fastest. Considering the quality and performance of smoothing operators, we have concluded that Tukey's bi-weight is the best operator among the

Table.1 Performance

	points/sec
MLS	2143
Lorentzian estimator	1150
Tukey's bi-weight estimator	1654
Smoothing distances	1789

four, because the quality of Figure 5(b) and 5(c) are almost same and Tukey's bi-weight is 40 % faster than the Lorentzian estimator.

3 Surface Fitting

3.1 Problems of Surface Extraction

When facilities are measured by laser scanners, the locations of scanners are often restricted and most components are measured only from one side. In addition, components may be partially hidden by other parts. Generally, the following cases are difficult to precisely calculate surface parameters:

- (1) Partially hidden components.
- (2) Dark-colored components.
- (3) Slim or short components.

Since dark-colored components absorb the energy of laser beams, returned signals become weak and noisy.

It is required to calculate shape parameters in such conditions as faithfully and precisely as possible.

3.2 Surface Fitting

For detecting surface parameters as faithfully and precisely as possible, we apply the fitting method proposed by Lukacs, et al. [14].

This method represents the distance function between surface $S(\mathbf{x} | \mathbf{a}) = 0$ and point \mathbf{p}_i as the following form:

$$d(\mathbf{a}, \mathbf{p}_i) = \sqrt{g(\mathbf{a}, \mathbf{p}_i)} - h(\mathbf{a}, \mathbf{p}_i), \quad (3.1)$$

where g and h are functions of parameters \mathbf{a} and the coordinate of \mathbf{p}_i . For example, the distance from a cylinder can be described as:

$$d(\mathbf{a}, \mathbf{p}_i) = \text{distance}(\text{axis}, \mathbf{p}_i) - r, \quad (3.2)$$

where *axis* is the straight line of the cylinder axis, and r is the radius of the cylinder.

In order to get rid of the square root of $d(\mathbf{a}, \mathbf{p}_i)^2$, Lukacs minimized :

$$\sum \tilde{d}(\mathbf{s}, \mathbf{p}_i)^2 = \sum \{(g - h^2) / 2h\}^2 \quad (3.2)$$

instead of $\sum (g - h^2)^2$, because the minimization of $\sum (g - h^2)^2$ amplifies the impact of noises and flattens the neighborhood of the solution. $\tilde{d}(\mathbf{s}, \mathbf{p}_i)$ has the same derivative values as $d(\mathbf{s}, \mathbf{p}_i)$ when $d(\mathbf{s}, \mathbf{p}_i) = \tilde{d}(\mathbf{s}, \mathbf{p}_i) = 0$. The minimization of this function tends to produce more stable and faithful fitting results than the ones of $\sum (g - h^2)^2$ and $d(\mathbf{a}, \mathbf{p}_i)^2$.

We implemented surface fitting methods for planes, spheres, cylinders, cones, and tori by minimizing Equation 3.2 using non-

linear optimization. Table 2 shows the number of parameters for each surface type. Generally, it is more difficult to precisely calculate surface parameters when the degree-of-freedom is larger. In addition, results are prone to noises when components are partially-hidden, dark-colored, short, or slim. In our experiments, the calculation of parameters often fails in such cases.

To solve this problem, we try to reduce the degree-of-freedom (DOF) by using the industrial standards and assembly constraints.

Table. 2: The DOF of primitive surfaces

plane	sphere	cylinder	cone	torus
3	4	5	6	7

3.3 Standards of Components

We maintain standard parameter lists and use them for surface fitting. For example, the diameters of pipes must be ones in Table 3 according to JIS standards.

When standards are incorporated, the feature-based modeling approach is useful. In our current implementation, the user first selects a feature type, which restricts parameter types and values, and then specifies a seed region for surface fitting by the region growing. Then the system grows the seed region and calculates surface parameters, which are selected as the nearest values in standard parameters. Finally, the system calculates surface parameters again by constraining the standard parameters to refine other unconstrained parameters. This approach is practically powerful for calculating surface equations precisely.

Table. 3: Diameters of pipes (JIS Standard)

10.5	13.8	17.3	21.7	27.2
34.0	42.7	48.6	60.5	76.3
89.1	101.6	114.3	139.8	165.2
190.7	216.3	267.4	318.5	355.6
406.4	457.2	500	508.8	558.8
...				

3.4 Assembly Constraints

Components are often connected to other components whose parameters are determined. Then we can reduce the number of parameters using assembly constraints. Assembly constraints include the following conditions:

- (1) The radii are the same. (DOF → - 1)
- (2) The axes are coaxial. (DOF → - 4)
- (3) The axes are parallel. (DOF → - 1)
- (4) The planar faces are on the same plane. (DOF → - 3)
- (5) The planar faces are parallel. (DOF → - 2)
- (6) Circle faces are coincident. (DOF: depends on types)

Since the minimization of Equation 3.3 is highly non-linear, assembly constraints make the calculation much more stable. The number of parameters can be further reduced by combining the industrial standards.

We implemented various minimization routines according to the types of assembly constraints. Figure 7 shows 3D models generated using constrained surface fitting. Figure (a)-(c) show examples that the system failed to calculate correct parameters when assembly constraints were not specified.

Figure 7(a) shows a partially-hidden cylinder, which was modelled using the same radius constraint. Figure 7(b) shows a coaxial cylinder. Since the height of this cylinder is short, the axis of the cylinder is prone to errors. Figure 7(c) shows a torus surface calculated using the same circle constraints. The calculation of tori is difficult in many cases because the DOF of tori is seven. Figure 7(d) shows 10 connected components.

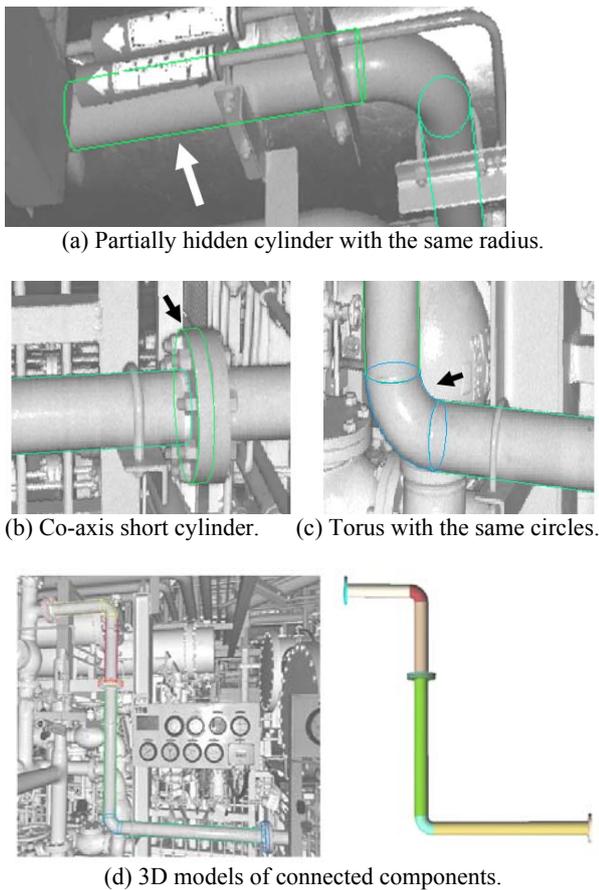


Figure 7: Experimental results of surface fitting

4 Conclusion

In this paper, we first discussed smoothing operators and compared the efficiency and performance of four types of operators. In our experiments, the surface estimator based on Tukey's bi-weight was the best among the four operators. Then we explained surface fitting methods based on the region growing and showed the industrial standards and assembly constraints were effective to calculate the surface equations of partially-hidden, dark-colored, short, or slim components.

In future work, we would like to automate surface detection and to support a rich set of features in our modeling system. In addition, we would like to improve the performance of smoothing operators.

Acknowledgements

The authors would like to thank Kenji Murakami for helping the development of the system. This research was partially supported by the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Scientific Research (B), 17360414 (2007) and 21360069 (2009).

References

- [1] TECHNICAL DATA IMAGER, [HTTP://WWW.ZF-LASER.COM](http://www.zf-laser.com), ZOLLER-FRÖHLICH.
- [2] SPECIFICATIONS TABLE, FARO PHOTON 120, [HTTP://LASER-SCANNER.FARO.COM](http://laser-scanner.faro.com), FARO.
- [3] VIEIRA, M. AND SHIMADA, K. 2005. SURFACE MESH SEGMENTATION AND SMOOTH SURFACE EXTRACTION THROUGH

- REGION GROWING. *COMPUTER AIDED GEOMETRIC DESIGN* 22, 771–792.
- [4] YANG, M. AND LEE, E. 1999. SEGMENTATION OF MEASURED POINT DATA USING A PARAMETRIC QUADRIC SURFACE APPROXIMATION. *COMPUTER-AIDED DESIGN* 31(7), 449–457.
- [5] MANGAN, A.P. AND WHITAKER, R.T. 1999. PARTITIONING 3D SURFACE MESHES USING WATERSHED SEGMENTATION. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS* 5(4), 308–321.
- [6] SHAMI, A. 2008. A SURVEY ON MESH SEGMENTATION TECHNIQUES. *COMPUTER GRAPHICS FORUM* 27(6), 1539–1556.
- [7] ALEXA, M., BEHR, J., COHEN-OR, D., FLEISHMAN, S., LEVIN, D. AND SILVA, C.T. 2001. POINT SET SURFACES. *IEEE VISUALIZATION*, 21–28.
- [8] LEVIN, D. 2003. MESH-INDEPENDENT SURFACE INTERPOLATION. *GEOMETRIC MODELING FOR SCIENTIFIC VISUALIZATION*, 37–49.
- [9] AMENTA, N. AND KIL, Y.J. 2004. DEFINING POINT-SET SURFACES. *ACM TRANSACTION ON GRAPHICS* 23(3), 264–270.
- [10] MASUDA, H. AND TANAKA, I. 2009. EXTRACTION OF SURFACE PRIMITIVES FROM NOISY LARGE-SCALE POINT-CLOUDS. *CAD&A*, 6(3), pp.387–398.
- [11] HUBER, P.J. 2003. *ROBUST STATISTICS*. WILEY SERIES IN PROBABILITY AND STATISTICS, WILEY-INTERSCIENCE.
- [12] STEWART, C.V. 1999. ROBUST PARAMETER ESTIMATION IN COMPUTER VISION. *SIAM REVIEW* 41(3), 513–537.
- [13] 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.
- [14] LUKACS, G., MARSHALL, A.D. AND MARTIN, R.R. 1998. FAITHFUL LEAST-SQUARES FITTING OF SPHERES, CYLINDERS, CONES AND TORI FOR RELIABLE SEGMENTATION. *PROCEEDINGS, 5TH EUROPEAN CONFERENCE ON COMPUTER VISION*, 671–686.