Simultaneous Entire Shape Registration of Multiple Depth Images Using Depth Difference and Shape Silhouette

Takuya Ushinohama, Yosuke Sawai, Satoshi Ono, and Hiroshi Kawasaki

Department of Information Science and Biomedical Engineering,
Graduate School of Science and Engineering, Kagoshima University, Japan

Abstract. This paper proposes a method of simultaneous global registration of multiple depth images which are obtained from multiple viewpoints. Unlike the previous method, the proposed method fully utilizes a silhouette-based cost function taking out-of-view and non-overlapping regions into account as well as depth differences at overlapping areas. With the combination of the above cost functions and a recent powerful meta-heuristics named self-adaptive Differential Evolution, it realizes the entire shape reconstruction from relatively small number (three or four) of depth images, which do not involve enough overlapping regions for Iterative Closest Point even if they are prealigned. In addition, to allow the technique to be applicable not only to time-of-flight sensors, but also projector-camera systems, which has deficient silhouette by occlusions, we propose a simple solution based on color-based silhouette. Experimental results show that the proposed method can reconstruct the entire shape only from three depth images of both synthetic and real data. The influence of noises and inaccurate silhouettes is also evaluated.

1 Introduction

3D shape measurement techniques have made significant progress, and have been widely used in various fields such as medical, educational, digital archiving and entertainment fields. For such purposes, range scanners are used to acquire 3D shapes of real-world objects and scenes. Since the 3D scanners can capture only one side of the object, multiple depth images should be captured from different viewpoints and aligned to recover the entire shape of the object as shown in Fig. 1; such process is called *registration* in the paper.

Registration techniques can be divided into two classes: global (coarse) and local (fine) registration. The purpose of global registration is to align the relative positions without prior knowledge of the initial clues until being possible to perform a local registration. The local registration requires a good prealignment to converge to the optimum solution. Iterative Closest Point (ICP) [1] and Simultaneous ICP [2] are well known methods for local registration. On the other hand, although a large number of studies have been conducted on global registration, general and robust algorithm has not been established yet.

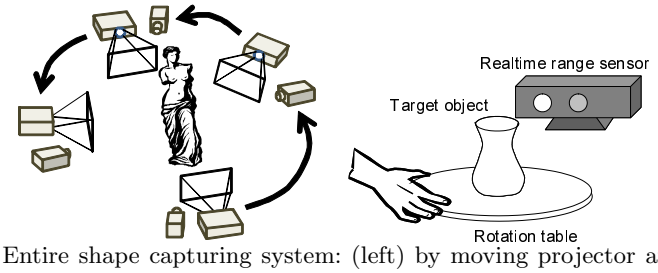


Fig. 1. Entire shape capturing system: (left) by moving projector and camera based 3D scanning systems and (right) by rotating the object.

Global registration methods are mainly categorized into two approaches: matching-based [3–5] and parameter-based methods [6–9]. Matching-based methods estimate an approximated position by utilizing 3D shape features [10]. With the methods, not only different types of algorithms are required depending on type of 3D shape, *e.g.*, mesh, depth image or point clouds, but also an appropriate 3D shape feature is necessary. Since 3D features cannot be retrieved stably due to changes on view-direction and scale and noise, the method usually requires manual support in practical cases. To solve the problem of the matching-based methods, parameter-based methods that use meta-heuristics have been proposed [7]. Parameter-based methods enable registration of the object regardless of type of 3D shape by direct pose-space search approach and it is also reported that the methods are robust against measurement environments. One severe drawback of the parameter-based methods is that it is generally a difficult task to find the optimum solution, because of vast search space, and thus, usually special assumptions are set, *e.g.*, an angle of rotation is limited [11], etc.

In this paper, we propose a global registration method with parameter-based approach which does not have special assumptions. To realize robust and practical convergence even if there are only small number of input, *i.e.*, three or four, we introduce a new cost function and optimizing technique as follows.

- Silhouette-based cost function, which takes out-of-view and non-overlapping regions into account as well as depth differences at overlapping areas. With the function, it only requires few overlapping regions which is not enough for pairwise registration approach [2].
- Simultaneous global registration method based on evolutionary computation algorithm named self-adaptive differential evolution (jDE) [12], which realizes registration of depth images without any prealignment.
- Color based silhouette for projector-camera system to compensate deficient silhouettes of depth image, which inevitably occurs by occlusions of stereo pair.

2 Related work

In order to improve the quality of registration which is prealigned, local registration method has been researched more then two decades. ICP and the extensions, which align the shape precisely by finding corresponding points as the closest point and minimizing the distance between them, are typical solutions [1, 2]. However, if the initial alignment was not enough accurate, the algorithm will fail because it easily makes wrong matches or it makes the cost to be 0 for the point which has no corresponding points nearby.

To solve the problem of initial alignment of ICP, many researches have been conducted known as global registration. Typical solution is a matching based approach using feature point [3–5]. Sample Consensus Initial Alignment (SACIA) [13] is a well known prealignment method that extracts key-points based on Fast Point Feature Histograms (FPFH), which is a fast variant of Point Feature Histograms (PFH) [14], and aligns two depth images by the key-point correspondences. Since those techniques are based on feature matching, it cannot be used if there are only small overlapping areas. Recently, the method which realizes registration even if there are only small overlapping areas is proposed [5]. This method aligns two wide baseline range scans by maximizing their *contour coherence*, that is, minimizing a distance between corresponding contours extracted from the range images. However, the method still require a certain amount of overlaps and minimum four scans are required.

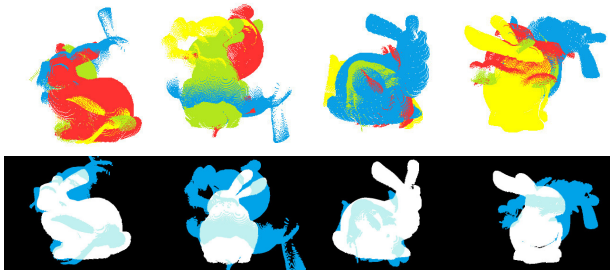
To solve the problem on matching based method, parameter-based methods have been proposed and those are summarized in [7]. Although previous methods have severe drawbacks on computational cost and vast local minima, our method overcomes the problem by using new cost functions with appropriate optimizing algorithm.

3 The proposed method

3.1 Overview

The proposed registration method first sets the specific viewpoint as the *source depth image*, transforms all the depth images to the viewpoint using the orientation and position parameters and merges them to make a *target depth image*. Then the target depth image is compared to the source depth image to calculate the cost based on silhouette as well as depth differences. Optimum parameters are found by minimizing the sum of the costs of different viewpoints.

As for the cost function of two depth images, usually depth differences of overlapping regions are used. With such method, if the number of images is small, size of overlapping region becomes small resulting in unstable registration. To overcome the problem, we propose a new cost function using non-overlapping region which is usually ignored for optimization. If all the depth images are transformed with the correct parameters, target depth image must be inside the silhouette of the source depth image. Based on the above standpoint, the proposed method simply puts cost on the part of the target depth image which



(a) View 1 (S_1: red) (b) View 2 (S_2: green) (c) View 3 (S_3: blue) (d) View 4 (S_4: yellow)

Fig. 2. Example of the source and the transformed targets (upper) and their silhouettes (lower). In the upper images, four models are classified by color. In the lower images, white, blue, and light blue represent source, target, and overlapped regions, respectively.

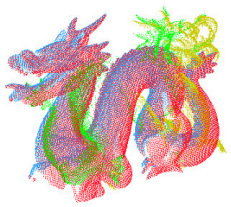


Fig. 3. Example of locally optimum, but not globally. Each pair, red-blue and green-yellow, are well aligned, but not between them.

is outside the source depth image. Fig. 2 shows examples of the non-overlapping regions; white, blue and light blue regions denote the source image, the target image which is not overlapped with source image, and the source image which is overlapped with the target image. We put constant cost on blue regions.

3.2 Acquiring multiple depth images around the object

In the method, we assume that the object is captured from multiple viewpoints to cover the entire shape. To achieve such scan, we assume either moving the scanner around the object or rotating the object with turn table, as shown in Fig. 1. Usually, to recover the entire shape using multiple scan data, it is required to scan the object from 8, 10 or more directions to have enough overlapping regions to stably run ICP or similar local registration methods. On the other hand, since our technique is based on silhouette of the object and we need almost no overlapping region in theory, just three or four scans are enough for entire shape registration, which greatly eases the scanning process and cannot be realized by previous methods; this is an important advantage of our method.

In terms of scanning device, since either object or scanner moves during the scan, we assume realtime 3D scanner which can capture the shape only with a short period of time. There are two types of realtime scanner, such as stereo based one, *e.g.*, video-projector [15] or laser based system [16], and Time-Of-Flight (TOF) based one, *e.g.*, Kinect2 [16] and Swiss ranger [17]. Stereo based scanner inevitably produces occluded regions because of wide base-line, whereas TOF sensor does not have such drawbacks. Since our method is based on silhouette of the object and its accuracy is affected by the silhouette, TOF scanner is more suitable in theory. However, depth accuracy and spatial resolution of realtime TOF sensor is basically lower than that of stereo based one, and thus,

stereo based scanner is considered main target in the paper. To compensate the occlusion problem, we propose an efficient method using color based silhouette which is captured by the camera of a projector-camera system.

4 Global registration of entire shape using silhouette

4.1 Cost function for simultaneous entire registration

A problem of simultaneous entire shape registration involves variables corresponding to rigid transformation parameters of all N depth images except the last one which was fixed to common world coordinate system. Therefore, solution vector $\boldsymbol{x}$ consists of set of rotation angles represented by quaternion $(\theta_1, \phi_{x1}, \phi_{y1}, \phi_{z1}), (\theta_2, \phi_{x2}, \phi_{y2}, \phi_{z2}), \ldots, (\theta_{N-1}, \phi_{xN-1}, \phi_{yN-1}, \phi_{zN-1})$ and set of translation vectors (x_1, y_1, z_1), $(x_2, y_2, z_2), \ldots, (x_{N-1}, y_{N-1}, z_{N-1})$.

In each calculation of objective function (cost function) for optimization, first, the proposed method renders all the silhouette images with depth information of reconstructed objects using $\boldsymbol{x}$ from all the viewpoints of depth sensors which capture the input shapes. Here, depth images which corresponds to each depth sensor are called *source depth images* and denoted as $S_1, \ldots, S_N$ and the rest of the depth images are merged for each viewpoints and called *target depth images* and denoted as $T_1, \ldots, T_N$. Fig. 2 shows examples of $S_1, \ldots, S_N$ (white) and $T_1, \ldots, T_N$ (blue and light blue) in the case of four scanned images.

The cost function proposed in this paper is calculated by comparing the target and source depth images. The cost function consists of the cost calculated by silhouettes and the cost calculated by depth differences. Since the proposed method does not depend entirely on the cost of overlapping region, the cost from silhouette is indispensable to prealignment. The cost from silhouette is based on the area size of the target regions stuck out from the silhouette of the source, and leads the target region to shrink into the view volume which is constructed by the source silhouette. It works efficiently to realize coarse, *i.e.*, globally consistent, registration at early stage of the optimization. The cost from depth difference at overlapping region works effectively on fine registration at the latter stage of the optimization; it makes the surface of source and the target close to each other. Based on the notation, objective function $F(\boldsymbol{x})$ is defined as follows:

$$F(\boldsymbol{x}) = \frac{1}{N} \sum_{k=1}^{N} f(T_k, S_k) \tag{1}$$

$$f(t, s) = \frac{1}{p} \sum_{i=1}^{p} \delta(t_i, s_i) \tag{2}$$

$$\delta(t_i, s_i) = \begin{cases} C_1 |s_i - t_i| & \text{(overlapping region)} \\ 1 & \text{(non-overlapping region)} \\ C_2 & \text{(outside of the view-field),} \end{cases} \tag{3}$$

where t and s denotes depth images of a target and a source, t_i and s_i denotes the depth values of pixel i in t and s, and p is the number of pixels in the depth

images. The cost from silhouette corresponds to the cost of non-overlapping region and outside of the view-field is represented by $\delta(t_i, s_i)$, which returns high penalty value C_1 when i is outside of the view-field, moderate value C_2 when the pixel in s_i exists in non-overlapping region, and a value based on the distance between s_i and t_i in overlapping region. Here, C_1 and C_2 should be adjusted in order that the cost of overlapping region can be almost the same or smaller than the cost of non-overlapping region; higher C_2 enhances the force to bring the source and the target close to each other, however, if it is too high, it entraps to local optima in which the distance between them is minimum. Reasonable values for them are [1, 10] and [1, 50] with our experience, respectively.

4.2 Global optimization by self-adaptive Differential Evolution

As described in Section 4.1, the entire registration problem involves many variables in $\boldsymbol{x}$ and the cost function $F(\boldsymbol{x})$ to be minimized. The least number of scans to reconstruct the entire shape is three, and in this case the problem consists of 21 variables. This optimization problem seems partially separable; variables for each transformation vector can basically be optimized individually. However, to escape from local optima, many variables should be changed simultaneously. The most hard local optima are that more than one pair (or group) of shapes are registered in each local coordinate systems respectively and the pairs (or group) are not located appropriately in the global coordinate system as shown in Fig. 3. To escape from such local optima, changes of a few variables are not sufficient; variables for two transformation vectors of two shapes should be changed simultaneously.

Therefore, the proposed registration method adopts meta-heuristics for global optimization named self-adaptive Differential Evolution (jDE)[12]. Differential Evolution (DE) [18, 19] is one of the most powerful stochastic real-parameter optimization algorithms. DE-variants and one other recent powerful EC algorithms named Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) [20] occupied high ranks in the standard numerical benchmarks such as IEEE Int'l Conf. Evolutionary Computation (CEC) competition on real parameter optimization. DE is suited to multimodal, separable problems, whereas CMA-ES to unimodal, non-separable ones.

jDE is simple but powerful self-adaptive DE algorithms that successfully eliminate control parameter adjustment by letting individuals have their own control parameter values and statistically changing the values. jDE showed the best performance in the competition "Evolutionary Computation in Dynamic and Uncertain Environments" in CEC2009 [21].

Unlike Genetic Algorithm [22], DE employs difference of solution candidates to explore the search space. DE generates solution candidates (vectors) by mutation and crossover. First, a mutant vector is generated as follows:

$$\boldsymbol{v}_{i,g+1} = \boldsymbol{x}_{r_1,g} + F_{i,g}(\boldsymbol{x}_{r_2,g} - \boldsymbol{x}_{r_3,g}), \tag{4}$$

where g denotes the index of generation. $\boldsymbol{x}_{r_1,g}$, $\boldsymbol{x}_{r_2,g}$, and $\boldsymbol{x}_{r_3,g}$ are randomly chosen vectors from the interval $[1, NP]$ ($r_1 \neq r_2 \neq r_3 \neq i$). Note that the

notations in this section are based on previous work [12], and have different meanings from the other sections of this paper. The above mutant vector are recombined with the target vector by crossover to produce trial vector $\boldsymbol{u}_{i,g+1}$ as follows:

$$u_{j,i,g+1} = \begin{cases} v_{j.i,g+1} \text{ if } \ rand_{j,i} \leq CR_{i,g} \text{ or } j = j_{rand} \\ x_{j,i,g} \quad \text{otherwise} \end{cases} \tag{5}$$

Finally, the better one of $\boldsymbol{u}_{i,g+1}$ and $\boldsymbol{x}_{i,g}$ survives and become a member of generation $g+1$ as follows:

$$\boldsymbol{x}_{i,g+1} = \begin{cases} \boldsymbol{u}_{i,g+1} \text{ if } f(\boldsymbol{u}_{i,g+1}) \leq f(\boldsymbol{x}_{i,g}) \\ \boldsymbol{x}_{i,g} \quad \text{otherwise} \end{cases} \tag{6}$$

Scale factor $F_{i,g}$ and crossover rate $CR_{i,g}$ are control parameters which deeply influence the search performance; for instance, small scale factor urges a local search, and low values of crossover rate are recommended for separable problems. In jDE, $F_{i,g}$ and $CR_{i,g}$ associated with i-th solution are statistically changed as follows:

$$F_{i,g+1} = \begin{cases} F_l + rand_1 \times F_u \text{ if } rand_2 < \tau_1 \\ F_{i,g} \qquad\qquad\quad \text{otherwise} \end{cases} \tag{7}$$

$$CR_{i,g+1} = \begin{cases} rand_3 \text{ if } rand_4 < \tau_2 \\ CR_{i,g} \text{ otherwise} \end{cases} \tag{8}$$

where F_l and F_u determine the range of the scale factor. Letting F_l and F_u be 0 and 1 respectively frees a human developer to adjust the control parameters except the population size. The probabilities to change $F_{i,g}$ and $CR_{i,g}$ are τ_1 and τ_2 respectively. They does not affect the result by choosing [0.05, 0.3], and are set to 0.1 according to the previous work [12]. By the above self adaptation process, adequate values of $F_{i,g}$ and $CR_{i,g}$ lead to better solution candidates that are more likely to survive, resulting in propagate those good parameter values.

4.3 Application to projector-camera system

When using a projector-camera system to obtain 3D shapes, self-occlusion causes deficiency of silhouette image due to stereo baseline. This silhouette deficiency may hamper the precise registration since the proposed method makes full use of silhouette for global optimization. Fig. 4(a) and (b) show a missing silhouettes obtained by a projector-camera system and a failed example of reconstructed shape with deficient silhouette images, respectively. With these silhouette images, the accurate reconstruction result cannot be evaluated properly; the target are not completely be covered by the deficient silhouettes, and the value of the cost function increases. The targets are then biased to make themselves gather into the center of the shape. The larger stereo baseline for the projector-camera system is used, the larger deficient regions the silhouette involves.

To solve the problem, when using a projector-camera system, we propose the method which compensates the deficient regions on silhouette images by

(a) Obtained silhouette image by projector-camera system. (b) Failed example reconstructed from (a) (c) Extracted region in color image. (d) Recovered silhouette.

Fig. 4. Example of silhouette recover with color image.

color information. To recover the deficient region of silhouette, we apply graph-cut algorithm using both color information and the silhouette of a depth image as the data cost. The extracted silhouette of the target object using graph-cut is shown in Fig.4(c). The comparison between the original silhouette and the recovered one using color information is shown by red regions in Fig.4(d). The above process is applied to each depth image. The cost function $\delta(t_i, s_i)$ in the recovered silhouette region in either source or target image is regarded as zero.

5 Evaluation

To verify the effectiveness of the proposed method, experiments were conducted with synthetic 3D model objects V_1, V_2, V_3, and V_4 shown in Fig. 5 and actual objects M_1, M_2, M_3, M_4, and M_5 measured by a projector-camera system. First, four experiments of different algorithm (Sec. 5.1), evaluation on influence of the size of the overlapped region (Sec. 5.2), comparison to previous work (Sec. 5.3), and noise influence (Sec. 5.4) were performed with the synthetic objects. Then, entire shape reconstructions using actual objects were demonstrated (Sec. 5.5). Finally, the influence of inaccurate silhouette (Section 5.6) and an ability on using multiple scans to recover a large object (Sec. 5.7) were examined.

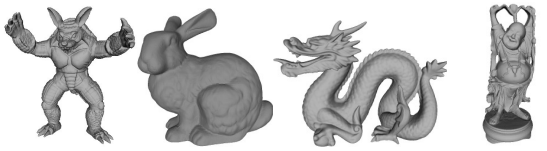
In the experiments with the synthetic objects, depth images were synthesized by a virtual TOF camera by rotating the object around y-axis. The target position to decide a view-direction of virtual sensor was set to the center of the object. The sensor's position was set behind the object center to z direction with distance d_z:

$$d_z = \frac{L_{max} + m}{\tan(fov)}, \tag{9}$$

where L_{max} is the maximum value between the height and width of the synthetic object and m denotes the margin set to be 1. The field of view fov of the camera was set to 30 degrees. The number of created points is from 10,000 to 25,000. The object size are normalized to be settled in the box region [-1.0, 1.0].

5.1 Comparison on optimization algorithms

In the first experiment, we compare the global optimization algorithm jDE with CMA-ES with restart (IPOP-CMA-ES) [23], that showed the best performance



(a) V_1 (b) V_2 (c) V_3 (d) V_4

Fig. 5. Synthetic 3D model objects used in the experiments.

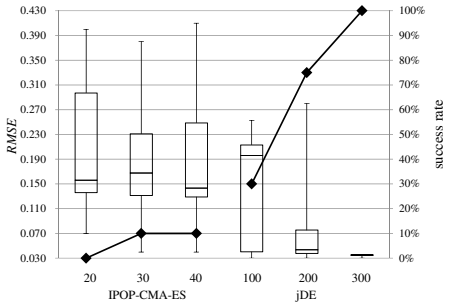


Fig. 6. Comparison on optimizers and their control parameters.

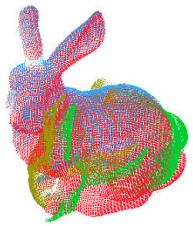


Fig. 7. Example local optimum from which IPOP-CMA-ES could not escape.

in Black Box Optimization Benchmarking 2012 [24]. Four depth images for each object V_1, V_2, V_3, and V_4 were used to reconstruct each model. According to the previous work [20, 12], their control parameters, λ in IPOP-CMA-ES and population size NP in jDE, were changed $20, 30, 40$ and $100, 200, 300$, respectively. IPOP-CMA-ES restarted and doubled the value of λ when no improvement of the best solution occurred during last 50, 80, or 100 generations, and terminated after three restarts. jDE stopped when no improvement of the best solution could be seen during 200, 300, or 500 generations. 20 independent runs have been performed for each object and for each algorithm and its control parameter configuration. C_1 and C_2 were set to 2 and 4, respectively. Ranges of variables for translation and rotation were [-1.0, 1.0] and [-180°, 180°], respectively. In each run, the entire shape registration was regarded as success when $RMSE$ was less than 0.05.

Fig. 6 demonstrates $RMSE$ (shown in box plots) and the success rates (shown by line graphs) of IPOP-CMA-ES and jDE in reconstructing V_2 from four images. The performance of jDE was better than IPOP-CMA-ES and improved in both the success rate and $RMSE$ as the population size increased, whereas the performance of IPOP-CMA-ES was worse than jDE and independent from λ. Since the fitness landscape of the registration problem based on the cost functions is globally multimodal in addition to local small perturbations, IPOP-CMA-ES was prone to fall into local optima as shown in Fig. 7. According

(a) eight images (45°) (b) four images (90°) (c) three images (120°)

Fig. 8. Difference of overlapped regions (red) on capturing angles.

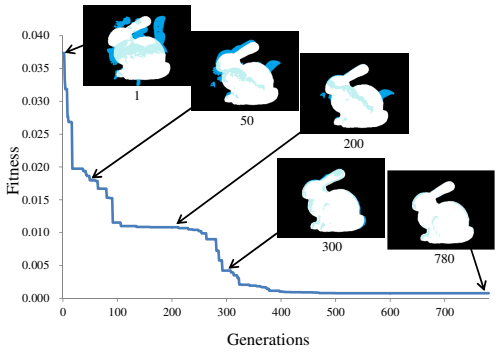


Fig. 9. Example transition of objective function value in reconstructing V_2 from three images.

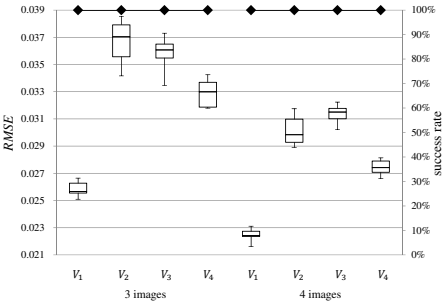


Fig. 10. Results on the different number of captured images.

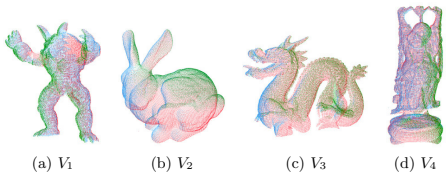
to the above results, we basically set NP to 300 in the later experiments in this paper.

5.2 Influence on size of overlapping region

Next, to show the advantage of the proposed method, we use only three or four images as the input to conduct registration; note that only little overlapped regions exist with such condition. We also test the registration with eight images which have enough overlapped region.

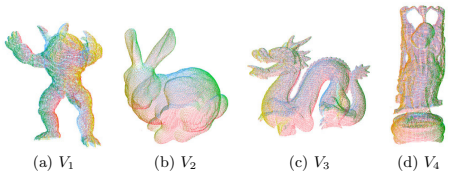
Input images were synthesized from virtual viewpoints with the interval of 120, 90 and 45 degrees, respectively. jDE stopped the search when no improvement of the best solution in the population could be seen in more than 1,000 generations. In the case reconstructing from eight images, NP was set to 1,600, a termination condition was extended to 4,000 generations, and only one run was performed due to the processing time. Fig. 8 shows the overlapped region with red color for each dataset. As the number of captured images decreased, the overlapped region significantly reduced so that the pair-wised methods could hardly align the images.

Fig. 9 shows an example transition of the objective function value of the best solution in reconstructing V_2 from three images. As the value decreased, the difference between the target and the source reduced. Fig. 10 demonstrates $RMSE$ shown in box plots and the success rates shown by line graphs, and



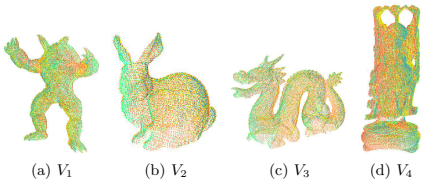
(a) V_1 (b) V_2 (c) V_3 (d) V_4

Fig. 11. Example results from three images by the proposed method.



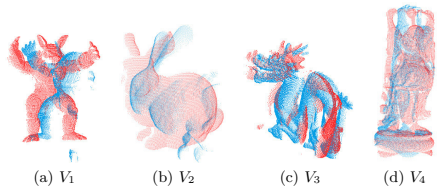
(a) V_1 (b) V_2 (c) V_3 (d) V_4

Fig. 12. Example results from four images by the proposed method.



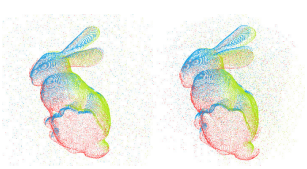
(a) V_1 (b) V_2 (c) V_3 (d) V_4

Fig. 13. Example results from eight images by the proposed method.



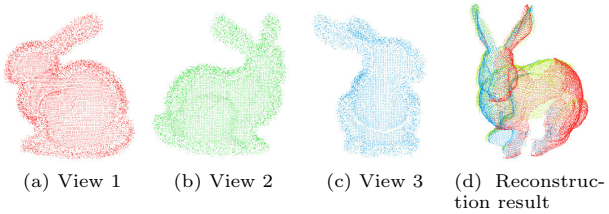
(a) V_1 (b) V_2 (c) V_3 (d) V_4

Fig. 14. Example pairwise registration results by SACIA.



(a) 10% 15%

Fig. 15. Reconstruction result from noisy depth images



(a) View 1 (b) View 2 (c) View 3 (d) Reconstruction result

Fig. 16. Reconstruction result from distorted shape images

Figs. 11, 12 and 13 show examples of the reconstructed 3D models. In the case reconstructing from eight images, the proposed method could succeed to align all images with RMSE of 4.55×10^{-2}, 6.63×10^{-2}, 6.40×10^{-2}, and 6.51×10^{-2} for V_1, V_2, V_3, and V_4, respectively. As shown in Fig. 10, the proposed method could reconstruct the entire shape from three or four images, despite the quite small overlapping regions.

5.3 Comparison with SACIA

The proposed method was compared with SACIA [13], a prealignment method based on key-point matching, in the case of using four images. Fig. 14 shows the examples of the pair-wised registration results by SACIA. Due to less overlapping regions, it was difficult for SACIA to align two depth images. In contrast, the proposed method could align just with four depth images as shown in Figs.10 and 12.

5.4 Robustness to noise and distortion

The robustness to noise of the proposed method was experimentally validated in this section with synthetic object V_2. First, the random uniform noise were added with the range of [-1.1, 1.1]. The amount of noise was 10% and 15% of the number of points of the tested objects [4]. Fig. 15 shows the results. The proposed method successfully reconstructed the entire shapes in noise levels 10%. In the case with 15% noise, the green target was a little off the other depth images, but being enough quality as prealignment.

Next, the influence of the measurement error and distortion was investigated by adding small perturbations to positions of all points with the radius of 0.001. Fig. 16 (a) to (c) show the object shapes whose points are perturbed, and Fig. 16(d) shows the reconstruction result with the obtained transformation parameters and with object shapes not involving the above point perturbations. From the results, we can confirm that the proposed method can avoid the influence of the small distortion and all depth images are successfully aligned.

5.5 Experiments with real data

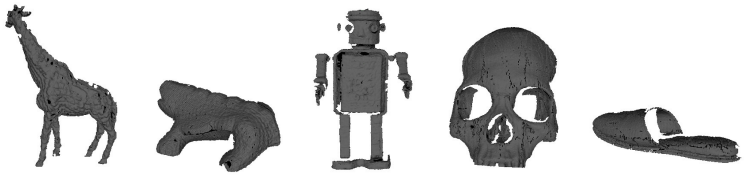
We use a projector-camera system to scan real objects. The system consists of EPSON EMP-1715 as a projector and Point Grey GRAS-20S4C-C as a camera. The baseline between the projector and the camera was 30 cm, and the distance between the object and the camera was 70 cm. The real object M_1, M_2, M_3, M_4, and M_5 were used. Each object was shot three times with rotating 120 degrees.

Figs. 18 and 19 show the reconstruction results of the proposed method without and with silhouette recovery, respectively; note that a cost function of *without* method is same as for ICP [2]. Without silhouette recovery, although shapes are gathered into the center of the shape, all of them cannot be reconstructed properly. To the contrary, the proposed method with silhouette recovery successfully reconstructed the entire shapes of all tested objects.

5.6 Influence of graph cut error in silhouette recovery

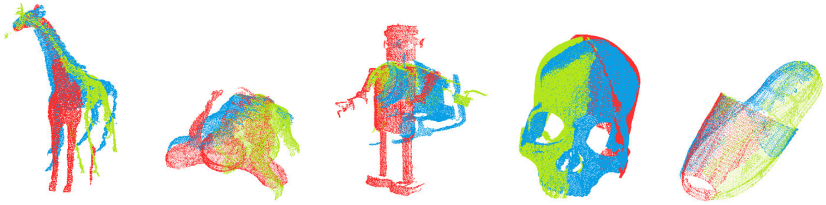
The influence of the error in graph cut for the silhouette recovery was investigated with M_2. The first case was the unexpected entering of the background to the silhouette. We assumed that one of the silhouette image extracted by graph cut involved the background region as shown in Fig 20 (a). The falsely detected region in the silhouette (the upper red-colored area of the source in Fig 20 (a)) has sufficient size to involve the target shape. The second case tested is the silhouette deficiency; a shaded area on the object was mistakenly recognized as background even after graph cut as shown in Fig.21(b).

Fig. 20 (b) and Fig. 21 (c) show the reconstruction results by the proposed method, demonstrating that the proposed method could reconstruct the entire shape with enough accuracy without being affected by the silhouette recovery failure.



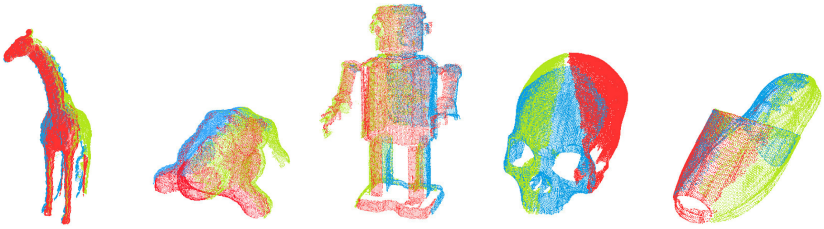
(a) M_1 (b) M_2 (c) M_3 (d) M_4 (e) M_5

Fig. 17. Real 3D model objects used in the experiments.



(a) M_1 (b) M_2 (c) M_3 (d) M_4 (e) M_5

Fig. 18. Reconstruction results without silhouette recovery.

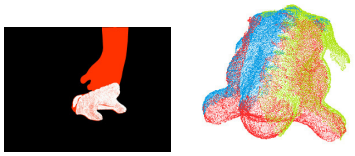


(a) M_1 (b) M_2 (c) M_3 (d) M_4 (e) M_5

Fig. 19. Reconstruction result with silhouette recovery.

5.7 Reconstruction of a large object from partially scanned data

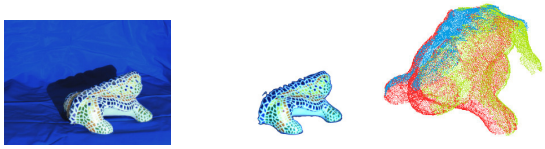
Reconstruction of a large object from partially scanned data, *e.g.*, face, hands, legs of statue, is practically important for wide applications [25]. Since our method basically assumes that the entire silhouette of object is available, preprocessing is necessary to apply our method, *i.e.*, constructing the entire silhouette of object by aligning multiple partial objects. To examine the proposed algorithm with the preprocessing, we conducted the experiment using synthetic data V_4. First, a virtual TOF camera scanned V_4 four times while moving the camera along y-axis in parallel. Then, as preprocessing, ICP was applied to the four images shown in Fig. 22(a), producing the depth image involving the entire silhouette of V_4 as shown in Fig. 22(b). Then, we do the same process by rotating the camera 120 degrees around y-axis. Finally, we apply the proposed method to the three preprocessed entire silhouettes and successfully reconstructed the entire shape as shown in Fig. 22(c).



(a) Silhouette involving background

(b) Reconstruct result from (a)

Fig. 20. Result from over-detected silhouette.

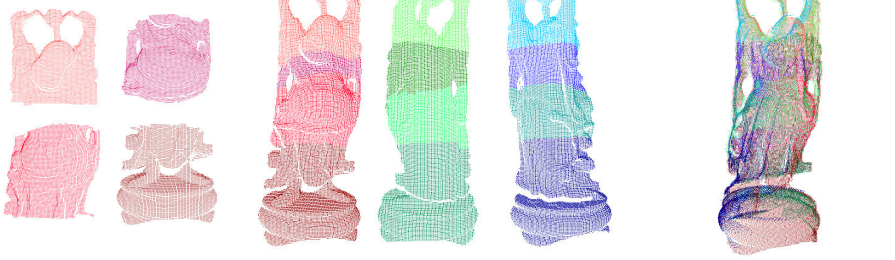


(a) Captured color image

(b) Graph cut result

(c) Reconstruct result

Fig. 21. Result from missing silhouette.



(a) Scanned image examples.

(b) Preprocessed images.

(c) Reconstruction result.

Fig. 22. Reconstruction result of large object with ICP-based preprocessing.

6 Conclusions

In this paper, we propose a simultaneous global registration method to achieve entire 3D shape reconstruction from relatively small number (three or four) of depth images. Even though input shapes do not involve enough overlapping regions, the proposed method reconstructs the entire shape without any initial positions. This is realized by using the cost which is assigned on out-of-view and non-overlapped regions as well as overlapped ones with global optimization method by jDE. In addition, our method is applicable to not only TOF sensors, but also projector-camera systems with color-combined silhouette recovery method. Experimental results showed that the proposed method can reconstruct entire 3D models even from three images captured by 120 degrees intervals which have almost no overlapping regions. Results using real data captured by a projector-camera system showed that color-based silhouette technique allows accurate reconstruction even if deficient in the depth images. Speeding up for optimization is our important future research.

References

1. J.Besl, P., D.McKay., N.: A method for registration of 3-d shapes. IEEE Trans. Pattern Analysis and Machine Intelligence **14** (1992) 239–256
2. NEUGEBAUER, P.: Geometrical cloning of 3d objects via simultaneous registration of multiple range image. Proc. Int'l Conf. Shape Modeling and Applications (1997) 130–139
3. Li, H., Hartley, R.: The 3d-3d registration problem revisited. In: Proc. Int'l Conf. Computer Vision. (2007) 1–8
4. Yang, J., Li, H., Jia, Y.: Go-icp: Solving 3d registration efficiently and globally optimally. In: IEEE Int'l Conf. Computer Vision. (2013) 1457–1464
5. Wang, R., Choi, J., Medioni, G.: 3d modeling from wide baseline range scans using contour coherence. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014) 4018–4025
6. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on **21** (1999) 433–449
7. Santamaría, J., Cordón, O., Damas, S.: A comparative study of state-of-the-art evolutionary image registration methods for 3d modeling. Computer Vision and Image Understanding **115** (2011) 1340–1354
8. Silva, L., Bellon, O.R.P., Boyer, K.: Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence **27** (2005) 762–776
9. Brunnstrom, K., Stoddart, A.J.: Genetic algorithms for free-form surface matching. In: Proc. Int'l Conf. Pattern Recognition. Volume 4. (1996) 689–693 vol.4
10. Salti, S., Tombari, F., Di Stefano, L.: A performance evaluation of 3d keypoint detectors. In: Int'l Conf. 3D Imaging, Modeling, Processing, Visualization and Transmission. (2011) 236–243
11. He, R., Narayana, P.A.: Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images. Computerized Medical Imaging and Graphics **26** (2002) 277–292
12. Brest, J., Greiner, S., Boskovic, B., Mernik, M., Zumer, V.: Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. Trans. Evolutionary Computation **10** (2006) 646–657
13. Rusu, R., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3d registration. In: Int'l Conf. Robotics and Automation. (2009) 3212–3217
14. Rusu, R., Blodow, N., Marton, Z., Beetz, M.: Aligning point cloud views using persistent feature histograms. In: Int'l Conf. Intelligent Robots and Systems. (2008) 3384–3391
15. Furukawa, R., Kawasaki, H.: Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning. In: IEEE Conf. 3DIM. (2005) 302–309
16. Zhang, Z.: Microsoft kinect sensor and its effect. MultiMedia **19** (2012) 4 –10
17. Mesa Imaging AG.: SwissRanger SR-4000 (2011) http://www.swissranger.ch/index.php.
18. Storn, R., Price, K.: Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optimization **11** (1997) 341–359
19. Das, S., Suganthan, P.N.: Differential evolution: A survey of the state-of-the-art. IEEE Trans. Evolutionary Computation **15** (2011) 4–31

20. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evol. Comput. **9** (2001) 159–195
21. Li, C., Yang, S., Nguyen, T.T., Yu, E.L., Yao, X., Jin, Y., g. Beyer, H., Suganthan, P.N.: Benchmark generator for cec' 2009 competition on dynamic optimization (2008)
22. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley, Reading (1989)
23. Auger, A., Hansen, N.: A restart cma evolution strategy with increasing population size. In: Congress Evolut. Comput. Volume 2. (2005) 1769–1776
24. Brockhoff, D., Auger, A., Hansen, N.: On the effect of mirroring in the IPOP active CMA-ES on the noiseless BBOB testbed. In: Proc. Ann. Conf. Genetic and Evolutionary Computation. (2012) 277–284
25. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The Digital Michelangelo Project: 3D scanning of large statues. In: Proceedings of ACM SIGGRAPH 2000. (2000) 131–144