An Improved ICP Algorithm Based on the Sensor Projection for Automatic 3D Registration

Sang-Hoon Kim[1] , Yong-Ho Hwang[1] , Hyun-Ki Hong[1] , and Min-Hyung Choi[2]

[1] Dept. of Image Eng., Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang Univ., 221 Huksuk-dong, Dongjak-ku, Seoul, 156-756, KOREA
ksh1974@imagelab.cau.ac.kr , hwangyongho@hotmail.com , honghk@cau.ac.kr

[2] Dept. of Computer Science and Eng., Univ. of Colorado at Denver, Colorado, USA
minchoi@acm.org

Abstract. Three-dimensional (3D) registration is the process aligning the range data sets form different views in a common coordinate system. In order to generate a complete 3D model, we need to refine the data sets after coarse registration. One of the most popular refinery techniques is the iterative closest point (ICP) algorithm, which starts with pre-estimated overlapping regions. This paper presents an improved ICP algorithm that can automatically register multiple 3D data sets from unknown viewpoints. The sensor projection that represents the mapping of the 3D data into its associated range image and a cross projection are used to determine the overlapping region of two range data sets. By combining ICP algorithm with the sensor projection, we can make an automatic registration of multiple 3D sets without pre-procedures that are prone to errors and any mechanical positioning device or manual assistance. The experimental results demonstrated that the proposed method can achieve more precise 3D registration of a couple of 3D data sets than previous methods.

1 Introduction

Range imagery is increasingly being used to model real objects and environments, and the laser sensors have simplified and automated the process of accurately measuring three-dimensional (3D) structure of a static environment [1]. Since it is not possible to scan an entire volumetric object at once due to topological and geometrical limitations, several range images showing only partial views of the object must be registered. Therefore, registration to align multiple 3D data sets in a common coordinate system is one of the most important problems in 3D data processing. For the registration process, each input data set consists of 3D points in the camera's local coordinate system. In order to register all input sets, a local coordinate of each 3D data set is transformed to a common coordinate, and the transformation between two data sets can be represented with a homography matrix. More specifically, the process provides a pose estimate of the input views that is a rigid body transformation with the six rotation and translation parameters. In general the relative sensor positions of several range sets can be estimated by mounting the sensor on a robot arm or keeping the sensor fixed and moving like an object on a turn-table [2].

The registration problem is composed of two phases: coarse registration and fine registration. In general, the coarse process obtains a rough estimation of 3D transforms by using mechanical positioning devices and manual processing. In order to refine the 3D estimate and make a complete 3D model, the fine registration is applied after coarse registration. The iterative closest point algorithm (ICP) is most widely used as the refinement method and calculates 3D rigid transformation of the closest points on the overlapping regions [3]. The two main difficulties in ICP, determining the extent of overlap in two scans and extending the method for multiple scans, have been a focus of the further research [4]. Namely, ICP requires *a priori* knowledge about an approximate estimation of the transformations, so starts with pre-estimated overlaps. Otherwise ICP tends to converge monotonically to the nearest local minimum of a mean square distance metric.

This paper presents an improved ICP algorithm that can automatically register multiple 3D data sets from unknown viewpoints. For a full automatic registration without an initial estimation process, we use the sensor projection matrix that is the mapping of the 3D data into its associated range image. The sensor projection matrix consists of the extrinsic and the intrinsic parameters as the camera projection does. The extrinsic parameters describe the position and the orientation of the sensor, and the intrinsic parameters contain measurements such as focal length, principal point, pixel aspect ratio and skew [5]. Since all range data is obtained with one range sensor, in general, the intrinsic parameters always remain unchanged. Then we use the covariance matrix to roughly obtain the extrinsic parameters that represent the relative 3D transformations between two inputs. The improved ICP method iteratively finds the closest point on a geometric entity to a given point on the overlapping regions based on the sensor projections. By combining ICP algorithm with the sensor projection constraint, we can solve the local minimum problem.

The remainder of the paper is organized as follows. In Sec. 2, previous studies are explained. In Sec. 3, an improved ICP algorithm is presented, and in Sec. 4, we demonstrate the experimental results and compare with previous methods. Finally, the conclusion is described in Sec. 5.

2 Previous studies

In the last few years, several algorithms for 3D registration have been proposed and can be classified into the semiautomatic and the automatic methods. Semiautomatic approaches require manual assistance including specification of initial pose estimates or rely on external pose measurement systems, so they have a couple of limitations and setting of equipments is needed [6]. For example, a mechanical positioning device can only deal with indoor-sized objects and a manual assistance may be inaccurate. On the contrary, automatic registration is to automatically recover the viewpoints from which the views were originally obtained without *a prior* knowledge about 3D transformation. The main constraint of most automatic methods is that many pre-processes including feature extraction, matching and surface segmentation are required. In order to calculate the pose for arbitrary rotation and translation parameters, we need to know at least 3 corresponding feature points between the 3D data sets. Once correspondences have been established, numerical minimization is used to determine the object's rotation and translation [7]. However, automatically detecting suitable features and matching them are very difficult, and currently no reliable methods exist. Furthermore, another approach is to ask the user to supply the features, but this is very labor intensive and often not very accurate.

From the viewpoints of constructing a complete 3D model, the registration problem is divided into the coarse registration and the fine. In the coarse process we use usually mechanical positioning devices or manual processing to obtain a rough estimate of 3D transforms. B. Horn proposed a closed-form solution to find the relationship between two coordinate systems of 3D points by using a unit quaternion from covariance matrix [8]. In addition, a refinery technique is needed to improve the 3D estimate and make a complete 3D model. After that P.J. Besl., *et. al.* presented ICP that optimizes 3D parameters based on Horn's method by using the closest points matching between two sets [3].

3 Proposed algorithm

This paper presents an improved ICP algorithm for automatic registration of multiple 3D data sets without *a prior* information about 3D transformations. The proposed iterative method uses the eigenvector of the covariance matrix and the sensor projection in an initial estimation. The eigenvector represents the direction of an object and defines a new axis at the centroid of the object. The analysis of the eigenvectors provides the relative sensor position in a common coordinate system. By using a cross projection based on the senor position, the overlapping regions can be detected. Finally, the improved ICP method iteratively finds the closest point on a geometric entity to a given point on the overlapping regions, and refines the sensor position.

3.1 Finding overlapping regions by cross projection

If an initial pose of the object differs so much from the real one, generally, it is difficult to construct a precise model due to self-occlusions. The more overlapping regions in 3D data sets we have, the more precise registration can be performed. Therefore, we assume that multiple-view range images have significant overlaps with each other [9]. By using an initial estimation of the relative sensor position and the sensor projection constraint, the proposed method finds the overlapping regions between two range data sets. The overlaps, which can be detected by two sensors at a time, are located in both 3D data sets. Fig. 1 shows overlapping regions on the first range data sets ($\boldsymbol{R}_1$) at the sensor $\boldsymbol{S}_1$ and the second ($\boldsymbol{R}_2$) at $\boldsymbol{S}_2$.

The overlapping regions on $\boldsymbol{R}_1$ are measured at the second sensor pose ($\boldsymbol{S}_2$), and those on $\boldsymbol{R}_2$ are also measured at the first ($\boldsymbol{S}_1$). So we can define the overlapping regions according to the visibility from two viewpoints. In order to find the overlapping regions, we propose a cross projection method that projects the first data set $\boldsymbol{R}_1$ into the second sensor position $\boldsymbol{S}_2$, and $\boldsymbol{R}_2$ into $\boldsymbol{S}_1$, respectively. Our method examines whether there are occlusions (scan errors) and self-occlusion. By analyzing the relation of the sensor direction vector and the vertex normal vector, we can find the overlapping regions. For example, when the angle between the sensor direction and the vertex normal vector ($\boldsymbol{V}_{\mathbf{n}}$) is lower than 90 degree, it is possible to scan the point by the sensor. Otherwise, we determine that the vertices are occluded, and those are not detected by the sensor.

$$(\boldsymbol{S}_2 - \boldsymbol{V}_1) \cdot \boldsymbol{V}_{\mathbf{n1}} > \mathbf{0}\text{: Overlap region in the } \boldsymbol{R}_1 \tag{1}$$

$$(\boldsymbol{S}_1 - \boldsymbol{V}_2) \cdot \boldsymbol{V}_{\mathbf{n1}} > \mathbf{0}\text{: Overlap region in the } \boldsymbol{R}_2\ , \tag{2}$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are vertex in $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, respectively. In short, the overlap in $\boldsymbol{R}_1$ is found by $\boldsymbol{S}_2$ and that in the $\boldsymbol{R}_2$ is by $\boldsymbol{S}_1$.

In the case of a concave object, there may be self-occlusion regions in the 3D data sets. As described in Fig. 2, the self-occlusion vertices are projected to the same pixel of the second range image by the second sensor projection. In this case, we examine the distance of the sensor position between vertices, and select the closest vertex from the sensor. The proposed cross projection can exclude the occlusion vertices and the self-occlusions, and find the overlapping regions between two views.

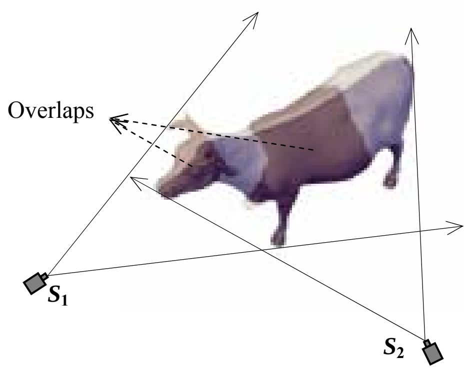


Fig. 1. Overlaps between two range sensors

3.2 Sensor projection matrix and sensor position

The sensor projection matrix is almost similar to the camera projection matrix. In general, the range sensor provides 3D range data sets (**X**) and the range image ($\boldsymbol{x}$) corresponding to the range data sets. We can compute easily the sensor projection ($\boldsymbol{P}$) and the sensor pose (**S**) using n corresponding points between range image and 3D range data sets [5]. The process is summarized as follows:

a. For each correspondence ($\boldsymbol{x}$ and **X**), $\mathbf{A_i}$ matrix (2 × 12) is computed.

b. Assemble n of $\mathbf{A_i}$ into a single **A** matrix (2n × 12).

c. Obtain the SVD (Singular Value Decomposition) of **A**. A unit singular vector corresponding to the smallest singular value is the solution $\boldsymbol{p}$. Specifically, if $\mathbf{A} = \mathbf{UDV}^{\mathbf{T}}$ with $\mathbf{D}$ diagonal with positive diagonal entries, arranged in descending order down the diagonal, then $\boldsymbol{p}$ is the last column of $\mathbf{V}$.

d. The $\boldsymbol{P}$ matrix is determined from $\boldsymbol{p}$, and the sensor pose ($\boldsymbol{S}$) is computed as follow:

$$\boldsymbol{S} = \boldsymbol{M}^{-1}\boldsymbol{p}_4 \ , \tag{3}$$

where $\boldsymbol{p}_4$ is the last column of $\boldsymbol{P}$, and $\boldsymbol{P} = \mathbf{M}[\ \mathbf{I}\ |\ \mathbf{M}^{-1}\mathbf{p}_4\ ] = \mathbf{KR}[\ \mathbf{I}\ |\ \mathbf{-S}\ ]$.

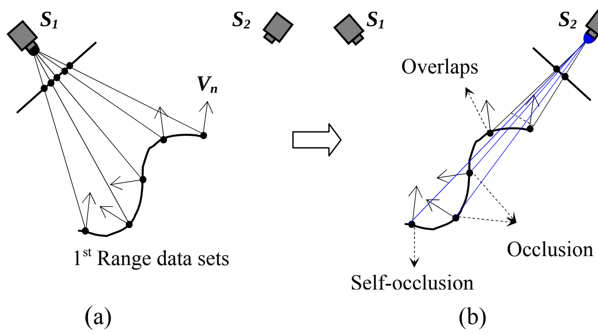


Fig. 2. The overlaps by cross-projection. (a) 1st range data from 1st sensor (b) 1st range data from 2nd sensor: the overlaps, occlusion and self-occlusion through the 2nd sensor and 1st range data

Since a range data is defined in its local coordinate system, all initial sensor positions obtained from multiple data sets are almost constant. For registration, we should find an accurate sensor pose in a common coordinate system. An accurate initial estimate can reduce the computation load and make the result of registration more reliable. By using a unit quaternion from closed-form covariance with the given corresponding point pairs in the two data sets, Horn suggested a method for registration of two different coordinate system [8]. Although our method is mathematically similar to Horn's, we do not use the corresponding point pairs, which are usually obtained in pre-processing or geometric limitation of 3D views. The proposed algorithm uses a similarity of 3D range data sets of an object from different views instead of feature correspondences. More specifically, we use the orthonormal (rotation) matrix that can be easily computed by eigenvalue decomposition of the covariance matrix instead of a unit quaternion. The eigenvector of the covariance matrix provides the major axis and the minor of 3D point clouds, so that it defines a new coordinate of the object. Three eigenvectors of the covariance matrix represent x, y, and z axes of the 3D data set, respectively. The obtained covariance matrix is used to define the object's local coordinate, and the centroid of the data sets is an origin of a new coordinate system.

In the first stage, the centroid ($\mathbf{C}$) of each range data sets is calculated as follows:

$$\mathbf{C} = \frac{1}{N}\sum_{j=0}^{N-1}\mathbf{V}_j \ , \tag{4}$$

where **V** and *N* represent 3D vertex in the range data sets and the number of vertices, respectively. In addition, we compute the covariance matrix (**Cov**) of each range data sets as follows:

$$\mathbf{Cov} = \frac{1}{\boldsymbol{N}} \sum_{j=0}^{N-1} (\mathbf{V}_j - \mathbf{C})(\mathbf{V}_j - \mathbf{C})^T \cdot \tag{5}$$

Let, $\mathbf{Cov_1}$ and $\mathbf{Cov_2}$ be covariance matrices of two range data sets ($\boldsymbol{R}_1$ and $\boldsymbol{R}_2$) respectively. We find two object coordinates by using eigenvalue decomposition of both covariance matrixes from $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

$$\begin{aligned} \mathbf{Cov_1} &= \mathbf{U_1 D_1 U_1}^{\mathbf{T}} \\ \mathbf{Cov_2} &= \mathbf{U_2 D_2 U_2}^{\mathbf{T}} \end{aligned} \quad , \tag{6}$$

where the diagonal matrices (**D**) and orthonormal matrices (**U**) represent eigenvalue and eigenvector of covariance matrices, then **U** provides a new object's coordinate. $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are defined again in the new coordinate by $\mathbf{U_1}$, $\mathbf{U_2}$ and the centroids of two range sets ($\mathbf{C_1}$, $\mathbf{C_2}$). In addition, a rigid transformation ($\boldsymbol{T}$) is found by $\mathbf{U_1}$ and $\mathbf{U_2}$, and an initial relative sensor position is approximated by the rigid transformation ($\boldsymbol{T}$).

$$\boldsymbol{T} = \begin{bmatrix} \mathbf{U}_1 \mathbf{U}_2^{-1} & \mathbf{C}_2 - \mathbf{C}_1 \\ 0_3^{\boldsymbol{T}} & 1 \end{bmatrix}. \tag{7}$$

3.3 3D registration by improved ICP algorithm

In general, ICP algorithm finds a rigid transformation to minimize the least-squared distance between the point pairs on the pre-determined overlapping regions. By using the sensor constraints and the cross projection, we can define the overlaps from the transformed data sets and the sensor positions, and then calculate 3D rigid transformation of the closest points on overlaps. More specifically, two range data sets are cross-projected into an initial position of the sensor, so an overlapping region is found. On the overlaps we find the closest point pairs, and calculate the transformations that can minimize the square distance metric between the points. The obtained transformations are used to optimize the initial sensor position for a more precise location. Our iterative method repeats the estimation of the sensor position and the detection of the overlapping regions. This process is repeated until the distance error value of closest point pair is minimized (Eq. 8), and we can optimize the sensor pose and 3D transformations of the range data.

$$\mathbf{E} = \sum_{i=1}^{n} \left\| \mathbf{V}_{1i} - \boldsymbol{R}(\mathbf{V}_{2i} - \mathbf{C}_2) - \boldsymbol{T} \right\|^2 , \tag{8}$$

where $\mathbf{V_1}$ and $\mathbf{V_2}$ are the closest point pairs of overlaps in two range sets and $\boldsymbol{C}_2$ is the centroid of $\mathbf{V_2}$. Rotation parameters ($\boldsymbol{R}$) is found by eigenvalue decomposition of two covariance matrices and translation ($\boldsymbol{T}$) is the displacement between the centroids of the points $\mathbf{V_1}$ and $\mathbf{V_2}$. Our method automatically finds the closest point pairs ($\mathbf{V_1}$ and

$\mathbf{V_2}$) on the overlapping regions between two 3D sets from unknown viewpoints. Fig. 3 provides the block diagram of the proposed algorithm

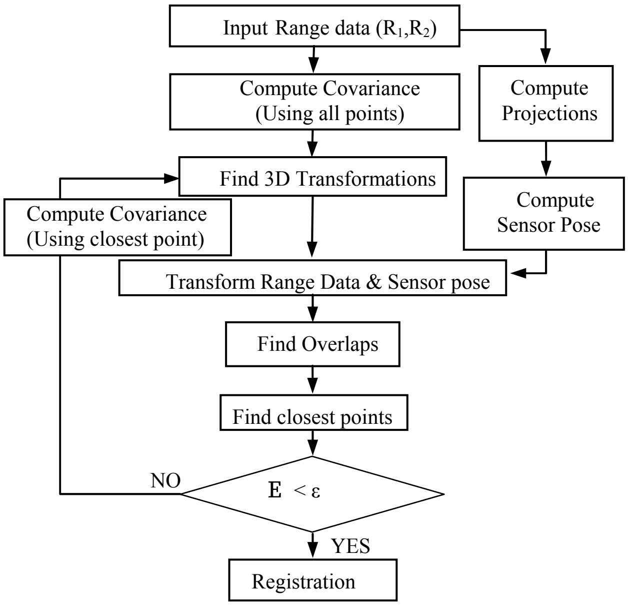


Fig. 3. Block diagram of the proposed method

4 Experimental results

We have demonstrated that the proposed algorithm can more precisely and efficiently register 3D data sets from unknown viewpoints than ICP method. The proposed algorithm has been tested on various 3D range images, which contain from 60K to 100K data points. The simulation is performed on PC with Intel Pentium 4 1.6GHz. In order to compare performance of each method, we use the virtual range scanning system and the real laser scanner with the calibrated turntable. We fixed the virtual sensor and acquired the "COW" range data sets by rotating the object with +60 degrees along Y-axis. "FACE" data is actually acquired by Minolta Vivid 700 laser scanner and the calibrated turntable with the same degrees. Consequently, we can compare precisely performance of the previous method with that of the proposed because 3D transformations are known.

Fig. 4 and 5 show the experimental results on two range data sets. As shown in Table 1, our method can make a precise registration without a precision milling machine or *a prior* 3D transformation between views. ICP algorithm computes 3D parameters of the closest points on the surface regions based on the sensor projection. On the contrary, the proposed method finds firstly the overlapping regions, and the points on these regions are considered. Therefore, the computation time for iteration in ICP is much longer than that in the proposed method. ICP converged to the local minimum of a mean square distance metric on "FACE" data, so the results on "FACE" data have much more errors than those on "COW". In the results by the proposed, the iteration times of "COW" is much less than those of "FACE", because "COW" has more obviously the major and the minor axis than "FACE". The results by the improved ICP show very small errors in each axis as shown in table 1, and these can be further vanished through an integration process. The proposed method obtains the major axis and the minor axis of 3D data sets to analyze the relative transformation, so it is difficult to register a totally spherical object.

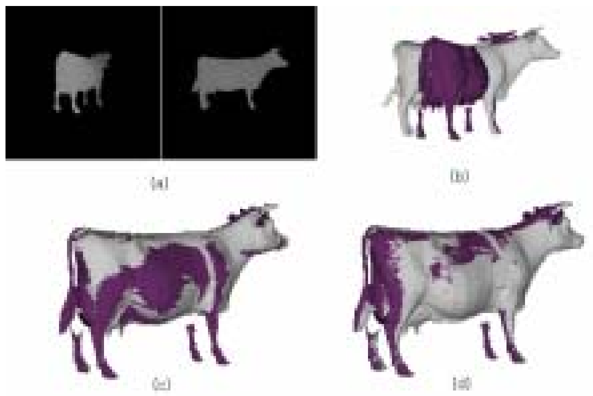


Fig. 4. Registration results of "COW" data sets. (a) Input range data sets (b) initial pose (c) registration by ICP (d) registration by improved ICP

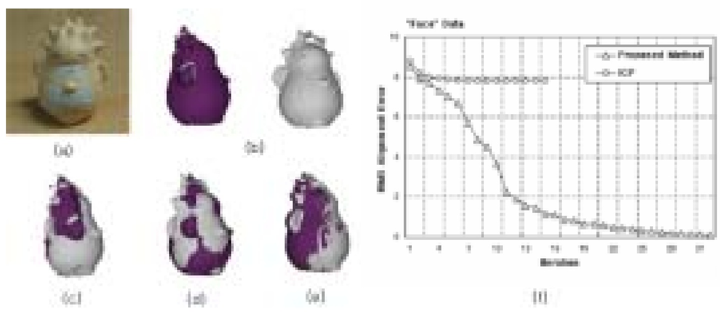
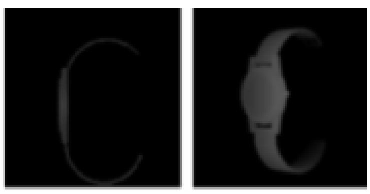


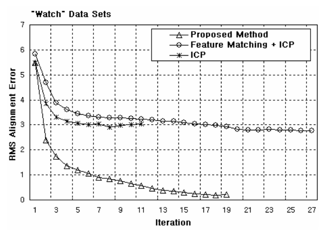
Fig. 5. Registration results of "FACE" data sets. (a) Real object (b) range data sets (c) initial pose (d) (e) registration by improved ICP (f) distance errors of the closest point pairs

Feature point extraction algorithm and ICP are generally combined to make an automatic registration of 3D data sets from unknown viewpoints. In this paper, the spin image is used as feature point extraction for registration [10]. Fig. 6 shows the comparison of the registration results by three methods. Because the overlaps between two data sets in "WATCH" are too small, it is difficult to find the corresponding points. As shown in Fig. 6 (a), only the proposed method accurately registered two data sets. The shape of "COPTER" has a bilateral symmetry, so the spin image is hard to establish the correspondence between two views. On the other hand, the proposed method computes the major and the minor axis of the object, and can cope with its symmetry in Fig. 6 (b). Since “DRIVER” has little overlap regions and its hilt is cylindrically symmetric, the spin image is impossible to match the points between views. The shapes of hilt in “DRIVER” from different viewpoints are almost alike, so our method is difficult to determine uniquely the minor axis of the hilt Fig. 6 (c) shows both methods failed to achieve a precise 3D registration.

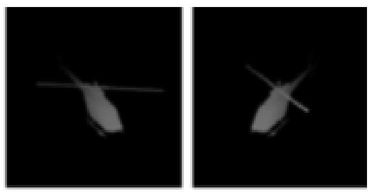
Table 1. The experimental results on two range data sets

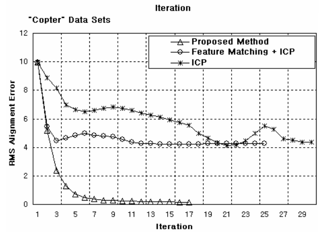
		Previous ICP	**Proposed Method**
COW	Iterations (times)	93	15
	CPU time (sec)	1867	87
	Rotation parameter Translation parameter	$\boldsymbol{\theta}_x$:-2.392, $\boldsymbol{\theta}_y$: 58.95, $\boldsymbol{\theta}_z$: 0.709 $\mathbf{T}_x$: -0.352, $\mathbf{T}_y$: 0.036, $\mathbf{T}_z$: 0.047	$\boldsymbol{\theta}_x$: -0.052, $\boldsymbol{\theta}_y$: 59.91, $\boldsymbol{\theta}_z$: 0.067 $\mathbf{T}_x$: 0.109, $\mathbf{T}_y$: -0.12, $\mathbf{T}_z$: 0.945
	Registration Error	$\boldsymbol{\theta}_x$: 2.392, $\boldsymbol{\theta}_y$: 1.051, $\boldsymbol{\theta}_z$: -0.709 $\mathbf{T}_x$: 0.352, $\mathbf{T}_y$: -0.036, $\mathbf{T}_z$: -0.047	$\boldsymbol{\theta}_x$: 0.052, $\boldsymbol{\theta}_y$: 0.09, $\boldsymbol{\theta}_z$: 0.067 $\mathbf{T}_x$: -0.109, $\mathbf{T}_y$: 0.12, $\mathbf{T}_z$:-0.945
FACE	Iterations (times)	92	87
	CPU time (sec)	694	123
	Rotation parameter Translation parameter	$\boldsymbol{\theta}_x$: 5.624, $\boldsymbol{\theta}_y$: 15.77, $\boldsymbol{\theta}_z$: 7.469 $\mathbf{T}_x$: -4.31, $\mathbf{T}_y$: 1.173, $\mathbf{T}_z$: 1.895	$\boldsymbol{\theta}_x$: 0.125, $\boldsymbol{\theta}_y$: 60.07, $\boldsymbol{\theta}_z$: -0.100 $\mathbf{T}_x$: 0.224, $\mathbf{T}_y$: -0.03, $\mathbf{T}_z$: -0.142
	Registration Error	$\boldsymbol{\theta}_x$: -5.624, $\boldsymbol{\theta}_y$: 44.23, $\boldsymbol{\theta}_z$: -7.469 $\mathbf{T}_x$: 4.31, $\mathbf{T}_y$: -1.173, $\mathbf{T}_z$: -1.895	$\boldsymbol{\theta}_x$: -0.125, $\boldsymbol{\theta}_y$: 0.07, $\boldsymbol{\theta}_z$: 0.100 $\mathbf{T}_x$: -0.224, $\mathbf{T}_y$: 0.03, $\mathbf{T}_z$: 0.142



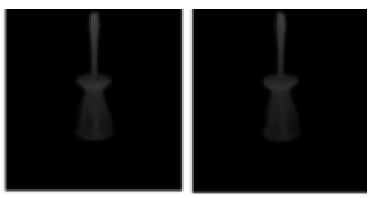


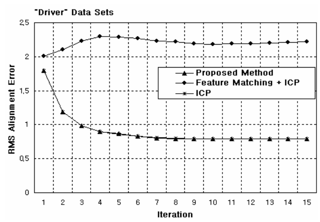
(a) WATCH





(b) COPTER





(c) DRIVER

Fig. 6. Comparison of previous methods and the proposed method

5 Conclusions

This paper presents an improved ICP algorithm that can automatically register multiple 3D data sets from unknown viewpoints without the preliminary processes including feature extraction and matching. The proposed method uses the sensor projection and the covariance matrix to estimate an initial position of the sensor. By using the cross projection based on the obtained senor position, we can find the overlapping regions. Finally, the improved ICP algorithm iteratively finds the closest point on a geometric entity to the given point on the overlapping regions, and refines the sensor position. The experimental results demonstrated that the proposed method can achieve a more precise 3D registration than previous methods. Further research includes a study on application to cylindrical or spherical objects. In addition, we will create photorealistic scene through 2D/3D alignment by using projections such as texture mapping.

Acknowledgment This research was supported by the Ministry of Education, Korea, and under the BK21 project, and the Ministry of Science and Technology, Korea, under the NRL(2000-N-NL-01-C-285) project.

References

1. L. Nyland, D. McAllister, V. Popescu, C. McCue, and A. Lastra, "The Impact of Dense Range Data on Computer Graphics," *In Proceedings of IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes* (1999) 3-10
2. D. F. Huber and M. Herbert, "Fully Automatic Registration of Multiple 3D Data Sets," *In Proceedings of IEEE Computer Society Workshop on Computer Vision Beyond the Visible Spectrum*, Dec. (2001)
3. P. J. Besl and N. D. Mckay, "A Method for Registration of 3-D Shapes," *IEEE Trans, Patt. Anal. Machine Intell.*, vol. 14, no. 2, Feb. (1992) 239-256
4. K. Pulli, "Multiview Registration for Large Data Sets," In Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling (1999) 160 – 168
5. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press (2000)
6. P. Neugebauer, "Geometrical Cloning of 3D Objects via Simultaneous Registration of Multiple range images," *In Proceedings of the 1997 International Conference on Shape Modeling and Applications*, Mar. (1997) 130 – 139
7. S. F. El-Hakim, P. Boulanger, F. Blais and J. A. Berladin, "A system for indoor 3-D mapping and virtual environments," *In Proceedings of Vidiometrics V (SPIE vol. 3174),* July (1997) 21-35
8. Horn, Berthold K. P., "Closed-Form Solution of Absolute Orientation Using Unit Quaternions," *Journal of the Optical Society of America*. A, vol. 4, no 4, April (1987) 629-642
9. G. Roth, "Registering Two Overlapping Range Image," *In Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling*(1999) 191-200
10. A. E. Johnson, M. H. Herbert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. on pattern analysis and machine intelligence*, vol.21, no.5, may (1999)