PE&RS, vol.67, no.10, pp.1143-1153, 2001

A ROBUST METHOD FOR REGISTERING GROUND-BASED LASER RANGE IMAGES OF URBAN OUTDOOR OBJECTS

Huijing ZHAO Ryosuke SHIBASAKI
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505 Japan
Email chou@skl.iis.u-tokyo.ac.jp

ABSTRACT

There is a potentially strong demand for detailed 3D spatial data of urban area. Ground-based laser range scanner is one of the promising devices to acquire range images of urban 3D objects. In this paper, the authors propose an automated registration method of multiple overlapping range images for the reconstruction of 3D urban objects. Registration is achieved in two steps, pair-wise registration and multiple registration assuming that one rotating axis of a laser range scanner is almost vertical. At first, pair-wise registration determines approximate values of four transformation parameters, a horizontal rotation angle and three translation parameters of a pair of neighboring range images. Then through multiple registration, the transformation parameters of all range images are adjusted using the approximate values so as to minimize the total errors. An outdoor experiment was conducted registering forty-two range images to construct a 3D model of a building in the campus of the Univ. of Tokyo. Accuracy of the model was examined using a 1:500 scale digital map and GPS measured location of viewpoints. Efficiency and Accuracy of the registration method is demonstrated in this paper.

INTRODUCTION

In a variety of applications ranging from visualization of urban landscape to advanced automobile/pedestrian guidance systems for ITS (intelligent transportation system), accurate and detailed 3D urban spatial databases are increasingly on demand. There are two approaches of 3D data acquisition - air-based and ground-based. Air-based data acquisition techniques, typically aerial survey, can cover relatively wide area, but usually fail to capture details of urban objects such as side walls (façade) of buildings. On the other hand, ground-based methods such as using vehicle-borne CCD cameras can easily cover such details of urban objects, though the spatial coverage may be limited. Recently, the reconstruction of 3D urban objects using ground-based techniques is attracting more attention because such details of urban objects, which can be easily viewed from streets or on ground surface, are found to be of importance in guidance system, for radio disturbance analysis in telecommunication and so forth. Many applications of 3D GIS involve user viewpoints on the ground, not in the air.

Several research groups in photogrammetry community invested in fusing the information from air and ground-based survey for the reconstruction of high-resolution building models within built-up area, where ground-level still images are pasted onto the building façades that are generated using air-based methods (e.g. Gruen 1998, Jaynes 1999). But an automatic calibration of ground-level view with aerial imagery is difficult to achieve, especially in highly dense urban areas.

Several researches using ground-based CCD cameras have demonstrated that 3D urban objects can be extracted using motion and stereo vision techniques (e.g. Ozawa *et.al.*1998). However, insufficient robustness in stereo matching, distortion from limited resolution and unstable geometry of CCD cameras are major obstacles to the operational uses of these methods.

Researches using range scanners are found in Thorpe *et.al.*1988 in a mobile navigation system; Kamgar-Parsi *et.al.*1991 in obtaining a map of ocean floor; Chen and Medioni 1992, Champleboux *et.al.*1992, Shum *et.al.*1994 in modeling small objects such as teeth, sculptures, mechanical parts, etc.; Ng *et.al.*1998 in constructing 3D models of indoor objects; Lemmens *et al.*1997, Haala *et al.*1998, Stilla and Jurkiewicz 1999 in air-borne systems. Through these research achievements, efficiency and accuracy of range scanners serving spatial data acquisition has been demonstrated. In addition, ground-based measurement using laser range scanner in urban environment has become technically feasible with the recent development of eye-safe laser range scanners. However, there is still no other researches addressing the reconstruction of 3D urban outdoor objects using ground-based laser range scanner.

Since one snapshot can not cover the entire 3D

urban object, one of the major challenges in applying ground-based laser range scanning for the reconstruction of 3D urban outdoor objects is the registration of multiple overlapping range images acquired at different locations (viewpoint) and with different viewing angles. This research is a contribution to the development of a robust method for registering a network of ground-based laser range images.

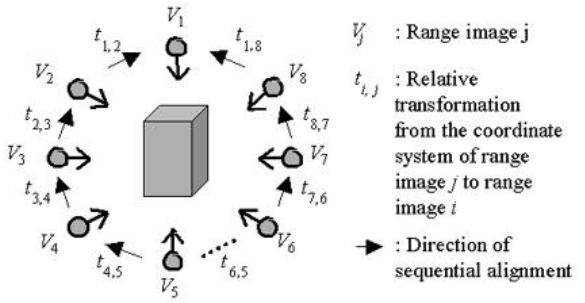

Figure 1. An example of range measurement from each façade of the building.

PROBLEM STATEMENT

Registering multiple range images – i.e., correctly aligning range images by transforming them into a common "global" coordinate system – is typically solved as a two-step procedure, pair-wise registration and multiple registration. Figure 1 shows a motivational example of measuring a building using a network of range images $\{V_1, V_2, \cdots, V_8\}$. If all range images are to be aligned to the coordinate system of V_1, while location and direction of viewpoints are unknown, the following steps are conducted. First, find the relative transformation matrixes $\{t_{ij} \mid (i, j) \in H\}$ between the coordinate systems of neighboring range images (pair-wise registration). It is always solved as a correspondence or matching problem. Next, find the absolute transformation matrixes $\{T_i \mid 1 \le i \le 8\}$ from each local coordinate system to the coordinate system of V_1 (multiple registration). In this step, $\{T_i \mid 1 \le i \le 8\}$ can be obtained by sequentially aligning $\{t_{ij} \mid (i, j) \in H\}$, however estimation errors in pair-wise registration might get accumulated. For example, V_6 can be aligned to the coordinate system of V_1 by $T_6 = t_{18} t_{87} t_{76}$, while V_5 be aligned to the coordinate system of V_1 by $T_5 = t_{1,2} t_{2,3} t_{3,4} t_{4,5}$. Then relative transformation matrix from the coordinate system of V_6 to V_5 is $t_{5,6}' = T_5^{-1} T_6 = t_{5,4} t_{4,3} t_{3,2} t_{2,1} t_{1,8} t_{8,7} t_{7,6}$. In practice, $t_{5,6}'$ is not equal to $t_{5,6}$ due to the accumulation of estimation error in $\{t_{ij} \mid (i, j) \in H\}$. Thus, the key problem that has to be tackled here is to minimize the accumulation of errors in pair-wise registration.

Pair-wise registration

Pair-wise registration has been at the core of many previous research efforts. Kamgar-Parsi *et al* 1991 matched the contours that extracted from different range images. Chen and Medioni, 1992 minimized the distances from control points of one view to the surfaces of another. Shum *et al.* 1994 exploited attribute graphs, which are generated using planar regions and their inter-relations. Krishnapuram and Casasent, 1989 determined the transformation parameters by extending the straight-line Hough Transform to three-dimensional space. Higuchi *et al.* 1995 converted the problem to the matching of two Spherical Attribute Images (SAI). Up to now, few works are addressed on urban outdoor area except the authors' previous research. In Zhao and Shibasaki, 1997, the authors presented a pair-wise registration method using planar faces, however robustness of the method is still insufficient to achieve full automation in some urban outdoor environment for the following reasons. First, range data in urban area are affected by many disturbances such as trees, window glasses, passing cars, pedestrians and so forth. Secondly, planar face far from the viewpoint is difficult to extract since the spatial resolution of range points on the planar face become lower. Thirdly, there are a lot of occlusions in individual range images, which block to identify corresponding planar faces. A registration method with robustness to high range noise and large number of irregular points is required.

In this research, we assume that laser range scanner is located in such a way that the horizontal rotating axis is vertical to the ground as shown in Figure 2. Based on this assumption, transformation parameters between range images are reduced from six – i.e. relative position (Δx, Δy, Δz) and three rotating angles (ω, ψ, κ) - to four (Δx, Δy, Δz and κ). A "Z-image" is introduced, which is generated by projecting range points onto a horizontal (X-Y) plane. Value of each pixel in Z-image is the number of range points falling into the pixel. See next section for a definition of "Z-image". Pair-wise registration is conducted in two steps, first matching Z-images to determine Δx, Δy and κ, then matching ground range points to estimate Δz.

Multiple registration

Many research efforts have focused on solving the error accumulation problem. Chen and Medioni, 1992 partially solved the problem by registering the newly added range image with the integrated range image consisting of all previously registered ones. Bergevin *et al.*1996 minimized the distance from a sequence of

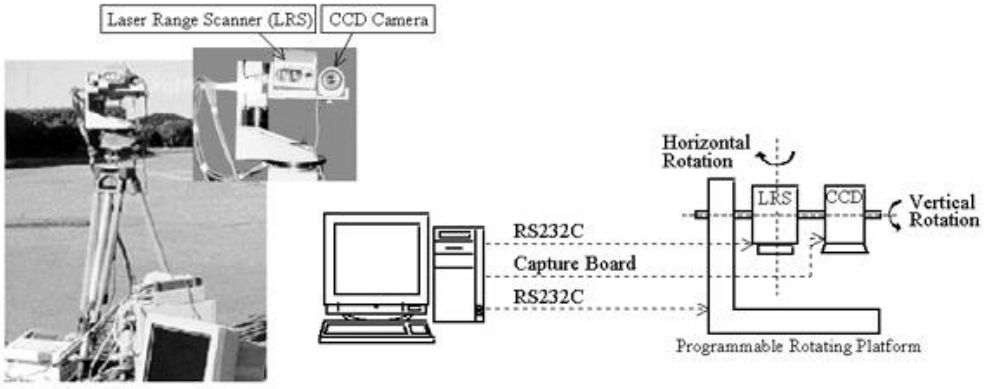

Figure 2. Architecture of the Laser Range Scanner.

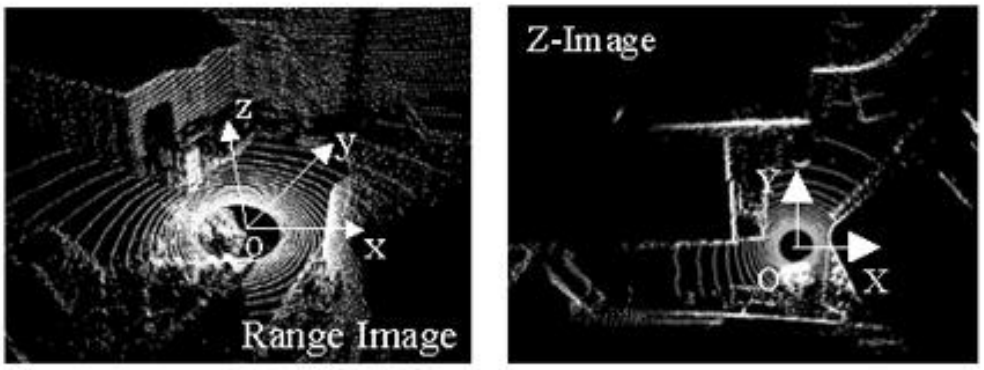

Figure 3. Generation of Z-Image from range image.

control points to the corresponding tangent planes in other range images as a least square problem. Shum *et al.* 1994 formulated the multiple registration as a problem of principal component analysis with missing data, where distance between corresponding planar faces are subject to minimization. B.Kamgar-Parsi *et al.* 1991 converted the problem to a resolution of conflicting situations that arise from the accumulation of pair-wise registration error.

In this research, we apply a simplified approach similar to B.Kamgar-Parsi *et al.* 1991 by minimizing the violation of absolute transformations obtained in multiple registration to the result of pair-wise registration as a weighted least square problem.

Outline of the paper

In the following sections, we will discuss the issues involved in both pair-wise and multiple registration. We present two experimental results in this paper. In the first experiment, registration of two range images is analyzed in detail to examine the methodological framework for pair-wise registration. In the second experiment, 42 overlapping range images are registered. Objective of the experiment is to examine the robustness of the pair-wise registration method, and test the accuracy and efficiency of multiple registration.

PAIR-WISE REGISTRATION USING Z-IMAGE

Definition of Z-image

A Z-image is introduced assuming that the horizontal rotating axis (Z-axis) of the laser range scanner is set vertical to the ground surface. It is generated by projecting range points onto a horizontal (X-Y) plane, where value of each pixel in Z-image is the number of range points falling into the pixel (see Figure 3). Vertical planar features like building surfaces are represented in Z-image as line segments, where a strong image feature in Z-image implies a high accumulation of range points along Z-axis. Comparing with the perspective view of range image, vertical building surfaces are emphasized in Z-image. On the other hand, trees, ground surface and other non-vertical planar features (e.g. a slope building wall) are weakened due to the low accumulation of range points along Z-axis. Strong linear features are extracted from Z-image for the purpose of pair-wise registration.

Matching Z-image

Matching Z-images is essentially a two-dimensional problem. Our method of matching Z-images can be generalized as follows.

1) Feature primitives

Line segments are exploited in Z-image matching, which are extracted using CFHT (*Curve Fitting Hough Transform*) (P.Liang, 1991).

2) Distance measure

We evaluate the similarity of the matching pairs by following the formalism defined in Boyer and Kak, 1988; Vosselman, 1992. In order to prevent mismatching of the poorly extracted features, we probabilistically evaluate the reliability of each line segment by exploiting the formalism defined in Kanatani, 1993.

3) Searching strategy

Searching for the best matching consists of two steps, coarse matching and fine adjustment. Coarse matching determines approximate transformation parameters between Z-images with an exhaustive search. In fine adjustment, transformation parameters are elaborated. Speed of convergence in the fine adjustment is improved by "strength" analysis.

In the following sections, we first discuss the reliability evaluation of line segments, then define the distance measure for the matching of Z-images. Searching strategy is addressed subsequently.

Reliability evaluation of line segments

The reliability definitions in this research follow and subsequently extend the formalism of Kanatani, 1993. Let D: $\{r_a \mid a = 1,..., N\}$ be a set of point measurements of line $l:(\bar{n},\bar{d})$ with a standard error ε. $\bar{n}$, $\bar{m}$ and $\bar{d}$ are line normal, directional vector and orthogonal distance respectively. Suppose r_a has its truth at $\bar{r}_a$, where $\Delta r_a = r_a - \bar{r}_a \sim N(0, \sigma^2)$. Let $l:(n,d)$ be the line parameters obtained by doing linear regression on D, $\Delta\theta$ be the small angle from $\bar{n}$ to n, $\Delta d = d - \bar{d}$. Then it has,

$$E[\Delta\theta] = O(\varepsilon^2) \tag{1}$$

$$E[\Delta d] = O(\varepsilon^2) \tag{2}$$

$$V[\Delta\theta] = \frac{N\sigma^2}{u} + O(\varepsilon^4) \stackrel{def}{=} \sigma_n^2 \tag{3}$$

$$V[\Delta d] = \frac{\sigma^2}{Nv} + O(\varepsilon^4) \stackrel{def}{=} \sigma_d^2 \tag{4}$$

where,

$$v = 1 - \frac{[\sum_{a=1}^{N}(\bar{m},\bar{r}_a)]^2}{N\sum_{a=1}^{N}(\bar{m},\bar{r}_a)^2} \tag{5}$$

$$u = N\sum_{a=1}^{N}(\bar{m},\bar{r}_a)^2 - [\sum_{a=1}^{N}(\bar{m},\bar{r}_a)]^2 \tag{6}$$

Reliability of parameter estimation on n and d is evaluated by σ_n and σ_d respectively. A larger variance of $\Delta\theta$ and Δd means less reliable parameter estimation of n and d.

Distance measure

Let $D=(P,L)$ be a structural description of Z-image. P denotes the group of image points, while L denotes the set of line segments extracted. Given two structural descriptions $D_1 = (P_1, L_1)$ and $D_2 = (P_2, L_2)$, and a mapping h from D_1 to D_2, distance from D_2 to D_1 can be defined using conditional information as follows (Boyer and Kak, 1988),

$$I_h(D_2 \mid D_1) = I_h(P_2 \mid P_1) + I_h(L_2 \mid L_1) \tag{7}$$

A. Conditional information of image points

Given a mapping h from D_2 to D_1, P_2 is divided into two groups.

1) $P_{out}^2 = \{r_{i_{out}}^2 \mid i_{out} = 1,..., N_{out}^2\}$ is the group of points having no match in P_1.
2) $P_{in}^2 = \{r_{i_{in}}^2 \mid i_{in} = 1,..., N_{in}^2\}$ is the group of point primitives having a matched point in P_1.

Conditional information of all image points can be formulated as follows,

$$I_h(P_2 \mid P_1) = I_h(P_{out}^2) + I_h(P_{in}^2 \mid P_1) \tag{8}$$

Let R be the maximal dimension of Z-Images, to describe a point primitive $r_i = (x_i, y_i)^T$ with resolution ε (pixels), it takes $2\log\frac{R}{\varepsilon}$ bits (Forstner, 1989; Vosselman, 1992). Then, the conditional information of image points without matched pairs can be calculated as follows,

$$I_h(P_{out}^2) = \sum_{i_{out}=1}^{N_{out}} I(r_{i_{out}}^2) = 2N_{out}\log\frac{R}{\varepsilon} \tag{9}$$

On the other hand, conditional information between the matched pair has been defined in Ingels, 1971; Blahut, 1987 using conditional probability as follows,

$$I(a_i \mid b_j) = -\log Prob(a_i \mid b_j) \tag{10}$$

Assuming that the conditional probability follows normal distribution, the conditional information of image points with matched pairs is calculated as follows,

$$I_h(P_{in}^2 \mid P_1) = \sum_{i_{in}=1,(j,i_{in})\in h}^{N_{in}} -\log\left(\frac{\varepsilon^2}{2\pi\sigma^2} e^{-\frac{(x_j^1 - h(x_{i_{in}}^2))^2 + (y_j^1 - h(y_{i_{in}}^2))^2}{2\sigma^2}}\right) \tag{11}$$

If r_j^1 and $r_{i_{in}}^2$ are truly matched, description of "matched" hypothesis for r_j^1 and $r_{i_{in}}^2$ should have fewer bits than a "no matched" hypothesis. Thus they must satisfy the following inequality,

$$-\log\frac{\varepsilon^2}{2\pi\sigma^2} e^{\frac{(x_j^1 - h(x_i^2))^2 + (y_j^1 - h(y_i^2))^2}{2\sigma^2}} \le 2\log\frac{R}{\varepsilon} \tag{12}$$

From the above discussion, it can be concluded that image point $r_j^1 = (x_j^1, y_j^1)^T$ and $r_i^2 = (x_i^2, y_i^2)^T$ are matched, if and only if it satisfies

$$\sqrt{(x_j^1 - h(x_i^2))^2 + (y_j^1 - h(y_i^2))^2} < t_r \tag{13}$$

where,

$$t_r = 2\sigma\sqrt{\ln\frac{R}{\sqrt{2\pi}\sigma}} \quad (14)$$

B. Conditional information of line segments

Given a mapping *h* from D_2 to D_1, L_2 is divided into two groups.

1) $L_{out}^2 = \{l_{i_{out}}^2 \mid i_{out} = 1,..., S_{out}^2\}$ is a set of line segments having no match in L_1.
2) $L_{in}^2 = \{l_{i_{in}}^2 \mid i_{in} = 1,..., S_{in}^2\}$ is a set of line segments having match in L_1.

Conditional information of all line segments can be formulated as follows,

$$I_h(L_2 \mid L_1) = I_h(L_{out}^2) + I_h(L_{in}^2 \mid L_1) \quad (15)$$

Line segments without matched pairs are represented using their point primitives as follows,

$$I_h(L_{out}^2) = 2\sum_{i_{out}=1}^{S_{out}} N_i \log\frac{R}{\varepsilon} \quad (16)$$

where, N_i is the number of image points in l_i.

Conditional information between two matched line segments consists of three components, match in image points, angle between line normal and difference between the orthogonal distance of the two line segments. It can be formulated as

$$I_h(l_i^2 \mid l_j^1) = I_h(p_i^2 \mid p_j^1) + I_h(angle(n_j^1, n_i^2)) + I_h(dis(d_j^1, d_i^2)) \quad (17)$$

Suppose Δq_s^k and Δd_s^k are the estimation error of n_s^k and d_s^k in l_s^k (*s=i,j; k=1,2*), following normal distribution, then it has

$$I_h(angle(n_j^1, n_i^2)) = -\log\left(\frac{\varepsilon}{\sqrt{2\pi}\sigma_{nij}} e^{-\frac{angle^2(l_j^1, h(l_i^2))}{2\sigma_{nij}^2}}\right) \quad (18)$$

$$I_h(dis(d_j^1, d_i^2)) = -\log\left(\frac{\varepsilon}{\sqrt{2\pi}\sigma_{dij}} e^{-\frac{dis^2(l_j^1, h(l_i^2))}{2\sigma_{dij}^2}}\right) \quad (19)$$

where σ_{nij} and σ_{dij} are the composite reliability of line normals and orthogonal distance respectively. They are denoted by

$$\sigma_{nij} = \sqrt{(\sigma_{nj}^1)^2 + (\sigma_{ni}^2)^2}$$

$$\sigma_{dij} = \sqrt{(\sigma_{dj}^1)^2 + (\sigma_{di}^2)^2}$$

If l_j^1 and l_i^2 are truly matched, description of matched hypothesis for l_j^1 and l_i^2 should have fewer bits than an unmatched hypothesis. Thus, we can conclude that line segments l_j^1 and l_i^2 are matched ones, if and only if it has

$$I_h(l_i^2 \mid l_j^1) < 2N_i^2 \log\frac{R}{\varepsilon} \stackrel{def}{=} t_i^2 \quad (20)$$

Searching for the best matching

A. Coarse matching

All line segment pairs with the same intersection angle or distance (if parallel), are selected as candidates in coarse matching. A matching yielding minimum distance measure is regarded as the best one, from which the initial approximate transformation between two images are determined. The initial transformation is denoted by T^0.

B. Fine adjustment

Fine adjustment can be formulated as an iterative minimization process, where in each iteration, the adjustment reducing the distance measure is made using gradient or Hessian matrix (Chen and Medioni, 1992; Champlebous *et al.* 1992). In case an objective function is complicated and its differentiation is not analytically easy like in our case, an alternative method is to directly compare the values of the objective function at slightly altered transformation parameter values to determine a direction of adjustment. However, this "hill climbing" method is rather time consuming.

We devise a method to achieve fine adjustment efficiently by analyzing the "strength" between matched line segment pairs. Line segments are selected because, first, line segments represent the intrinstric structure of Z-images, which might help in overcoming local optimum caused by image points; secondly, reliable line segment pairs will have less conditional information than unreliable ones, which might help in discriminating the contributions from different line segment pairs, and avoiding the mismatching due to unreliable line segments.

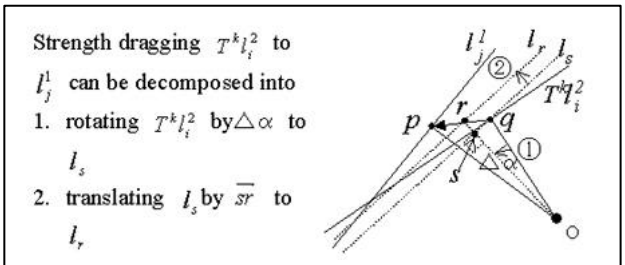

Figure 4. Definition of strength from $T^k l_i^2$ to l_j^1 in the kth iteration.

Strength between a matched line segment pair l_j^1 and l_i^2 is defined as follows. Let T^k be the revised transformation in the *kth* iteration. Given a step value $\triangle\alpha$ of the rotation angle, the strength ΔT_{ij}^k dragging $T^k l_i^2$ to l_j^1 can be decomposed into first rotating $T^k l_i^2$ by $\triangle\alpha$ to l_s ($\stackrel{def}{=} \triangle R_{ij}^k$), then translating l_s by $\overline{sr}$ to

$l_r \left(\stackrel{def}{=} \triangle Sh_{ij}^k\right)$ (Figure 4). Weight of the strength ΔT_{ij}^k is defined by using the conditional information between l_i^2 and l_j^1 as follows,

$$w_{ij}^k = I_{T^k}(l_i^2 \mid l_j^1) - I_{T^k \circ \Delta T_{ij}^k}(l_i^2 \mid l_j^1) \quad (21)$$

Adjustment $\triangle T^k$ in the *kth* iteration is obtained using the weighed average of the strength of all matched line segment pairs. It is defined as follows.

$$\Delta T^k = \Delta R^k \circ \Delta Sh^k \quad (22)$$

where,

$$\Delta R^k = \frac{1}{w} \sum_{(i,j)\in h} (w_{ij}^k * \Delta R_{ij}^k)$$

$$\Delta Sh^k = \frac{1}{w} \sum_{(i,j)\in h} (w_{ij}^k * \Delta Sh_{ij}^k)$$

$$w = \sum_{(i,j)\in h} w_{ij}^k$$

However, the claimed adjustment by strength analysis may fail to reduce the distance measure defined in (7). In that case, the hill climbing method is applied. The iteration stops when reduction rate of the distance measure becomes lower than a threshold value.

Matching ground range points

With the registration methods described so far, horizontal rotation angle and shift vector (Δx, Δy) are determined. Determination of the translation parameter along Z-axis is conducted in two steps. An initial configuration of Δz is decided by minimizing the distance between the ground range points of two range images. A fine adjustment of the initial Δz is conducted using *iterated closest-point (ICP)* algorithm (Besl and McKay 1992). An initial configuration of Δz is calculated as follows.

1. Transforming a pair of range images (image1 and image2) horizontally using the transformation parameters determined by the Z-image matching, and projecting each range image onto a common horizontal (X-Y) plane.
2. Tessellating the horizontal plane into grid cells, height of the ground surface at each grid cell is estimated as the minimum Z value of the range points within the grid cell. In addition, the ground height in the grid cells near the laser range scanner is interpolated, because range points can not be acquired due to physical limitation of the scanner. The estimation of ground height is conducted using the range image1 and 2 respectively, which results in two data sets of grid-based ground height.
3. At each grid cell, the estimated ground height values from the image1 and 2 are compared. If they are equivalent within a given threshold value, the grid cell is regarded as matched. The translation parameter along Z-axis is determined by maximizing the number of the matched grid cells.

MULTIPLE REGISTRATION

Suppose $\{t_{ij} \mid (i,j) \in H\}$ is the set of inter-frame relations obtained in pair-wise registration, where t_{ij} is the relative transformation matrix from the coordinate system of range image #*j* to range image #*i*. Let T_i be the absolute transformation matrix from the coordinate system of range image #*i* to a global coordinate system. Let $\bar{t}_{ij} = T_i^{-1} \circ T_j$ be the relative transformation matrix obtained in multiple registration. In this research, we first do sequential alignment to set initial values to $\{T_i\}$, then perform a simultaneous adjustment to reduce the differences between $\{t_{ij}\}$ and $\{\bar{t}_{ij}\}$. Difference between t_{ij} and $\bar{t}_{ij}$ is evaluated using the following three parameters, d_{ij}, b_{ijk} and q_{ij}, as shown in Figure 5. Simultaneous adjustment is formulated as a least square minimization of the following cost function.

$$E_g = w_d \sum_{(i,j)\in h} (d_{ij} - \bar{d}_{ij})^2 + w_b \sum_{(i,j)\in h;(i,k)\in h} (b_{ijk} - \bar{b}_{ijk})^2 + w_q \sum_{(i,j)\in h} (q_{ij} - \bar{q}_{ij})^2 \quad (23)$$

Where, $(d_{ij}, \boldsymbol{b}_{ijk}, q_{ij})$ denote the parameters determined in pair-wise reigstration, $(\bar{d}_{ij}, \bar{\boldsymbol{b}}_{ijk}, \bar{\boldsymbol{q}}_{ij})$ denote those obtained in multiple registration. (w_d, w_b, w_q) serve as weights for different contributions of the components, which is defined using training values.

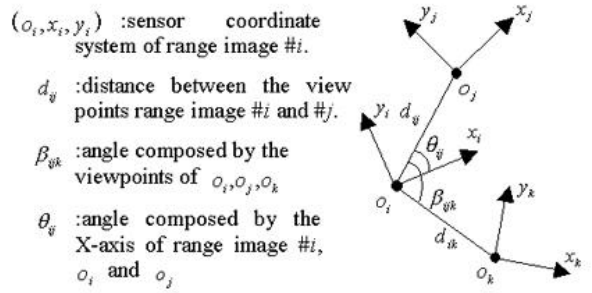

Figure 5. Parameters in multiple registration.

On the other hand, let $\{\bar{p}_k\}$ be the estimated viewpoints in multiple registration. If the absolute positions $\{p_k\}$, $k \in K$ in world coordinate system of more than three viewpoints are measured, e.g. using GPS, A^0 is the *affine* transformation matrix, with $\min \sum_{k\in K} (A^0 \bar{p}_k - p_k)^2$, then all range images can be aligned to the world coordinate system by $T_i^0 = A^0 \circ T_i$. If the value of A^0 is set by sequential alignment, in simultaneous adjustment, it is also required to minimize the residuals between $\{A^0 \bar{p}_k\}$ and $\{p_k\}$, $k \in K$. Hence, the cost function of least square minimization

becomes

$$E_g = w_d \sum_{(i,j)\in h} (d_{ij} - \bar{d}_{ij})^2 + w_b \sum_{(i,j)\in h;(i,k)\in h} (b_{ijk} - \bar{b}_{ijk})^2 + w_q \sum_{(i,j)\in h} (q_{ij} - \bar{q}_{ij})^2 + w_p \sum_{k\in K} \left\| p_k - A^0 \bar{p}_k \right\| \quad (24)$$

EXPERIMENT RESULTS AND DISCUSSIONS

In this section, we present the results of two outdoor experiments. The first experiment is to test the validaty of the method for pair-wise registration using Z-image. The second experiment aims at examining the robustness of pair-wise registration, and test the efficiency and accuracy of the multiple registration, where 42 overlapping range images are registered to construct a 3D model of the buildings in the campus of University of Tokyo. Range images used in the experiment are measured by [-180°,+180°] in horizontal rotation angle and [-20°,+40°] in vertical rotation angle, with the resolution of 2 samples per horizontal degree and 1 sample per vertical degree. Laser range measurement has an accuracy of ±5cm. Two sets of data are used as the ground truth to examine the registration accuracy. They are 1) a 1:500 scale digital map and, 2) coordinates of viewpoint locations of the laser range scanner measured by GPS with an accuracy of ±20cm.

Experiment of pair-wise registration

Figure 6 shows two range images, and the Z-images generated with a resolution of 0.25m/pixel. Reliability evaluation results of line segments in Z-image1 and 2 are listed in Table 1. From Table 1, it is obvious that line segments of #0~#3 in both images have rather high reliability, while #6,#7 in Z-image1 and #4,#6,#8,#9 in Z-image2 are wrong extractions. In Table 1, we can find that unreliable line segments always have larger values of s_n or s_d than reliable ones, although there's no obvious difference in variance of regression residuals between the reliable and unreliable line segments. This

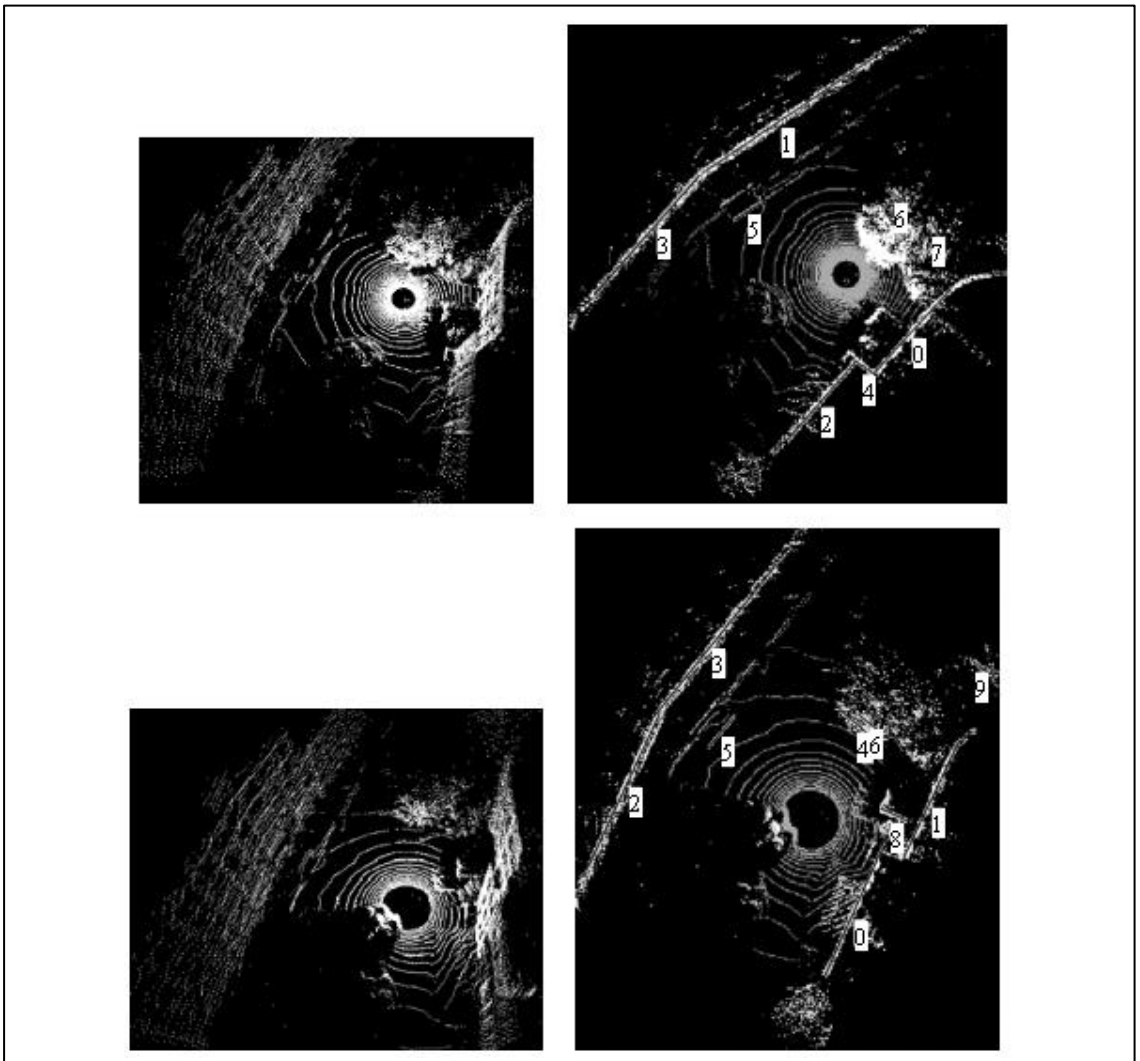

Figure 6. Range images, Z-Images and extracted of line segments (a) range image 1 (b) Z-Image 1 (c) range image 2 (d) Z-Image 2 *(The sequential numbers of line segments are consistent with those in Table 1)*

suggest s_n and s_d are more appropriate to characterize the reliability of line parameter estimation.

Table 1. Reliability evaluation of line segments extraction.
Res.Var. : variance of regression residuals;
Length : length of the line segment;
Point Number : number of the image points used in line parameter estimation.

Z-Image 1					
Line No.	s_n (°)	s_d (pixel)	Res.Var. (pixel)	Length (pixel)	Point Number
0	0.08	0.502	1.043	83	1362
1	0.066	0.452	1.441	169	899
2	0.126	0.635	1.05	78	561
3	0.168	0.86	1.7	101	453
4	0.93	1.82	0.882	16	202
5	0.471	2.723	0.253	25	20
6	2.221	6.23	2.094	43	19
7	4.097	9.875	1.338	47	43
Z-Image 2					
Line No.	s_n (°)	s_d (pixel)	Res.Var. (pixel)	Length (pixel)	Point Number
0	0.041	0.248	0.777	93	1814
1	0.063	0.451	0.769	87	1237
2	0.075	0.464	1.351	129	859
3	0.1	0.734	1.638	129	767
4	1.163	7.321	0.452	9	85
5	2.3	5.198	1.548	14	66
6	3.524	20.462	2.092	16	84
7	2.772	10.004	1.039	10	79
8	2.93	12.836	0.934	13	30
9	1.712	1.809	2.419	35	44

In coarse matching, given a threshold value for the maximum difference in intersection angle (=2°), and the maximum difference in distance (=2pixel) respectively, a total of 48 candidates are obtained. Among all the candidates, line segment pairs (#2, #0), (#0, #1), (#3, #2), (#1, #3), (#5, #5) of Z-image1 and 2 are selected and transformation T^0 is determined. Conditional information and strength between the matched line segment pairs is listed in Table 2. The strength is computed with a step value of 0.1 ° rotation angle. From Table 2, we can find that the conditional information and the weight of strength between line segment pair (#5, #5) (I_{ij}=186, V_{ij}^0=7.0192) is much lower than the others, whereas it has larger values of s_{nij} (=0.587) and s_{dij} (=3.584). Large value of s_{nij} or s_{dij} means poor reliability on parameter estimation of line normal or orthogonal distance, while low weight of strength means a smaller contribution to the adjustment of transformation. Thus, by the conditional information and the strength analysis we can discriminate effectively between reliable and unreliable line segment pairs and their contributions to the adjustment.

Table 2. Conditional information and strength between matched line segment pairs that obtained in coarse matching

Matched Pair *(j, i)*	q_{ij} (°)	d_{ij} (pixel)	s_{nij} (°)	s_{dij} (pixel)	I_{ij} (bits)	V_{ij}^0 (bits)
(2,0)	0.178	2.853	0.132	0.682	10731	1213.4
(0,1)	1.07	2.685	0.102	0.675	7799	376.91
(3,2)	1.014	2.423	0.184	0.977	5442	-8.247
(1,3)	0.178	2.159	0.12	0.862	4774	405.8
(5,5)	1.922	3.722	0.587	3.584	186	7.0192

In Table 2, we can find that there is a minus weight of strength for line segment pair (#3, #2). This means that the adjustment claimed by the strength analysis for line segment pair (#3, #2) failed in reducing the conditional information between them, since conditional information evaluate not only the distance between line parameters but also the matching fitness between line points, which is omitted in the strength analysis. A minus weight of strength will be ignored in calculating the weighed average of strength. Figure 7 shows the change of distance measure in fine adjustment and the method used to decide the adjustment in each iteration. The strength analysis is found efficient for adjustment, as it requires approximately one-thirtieth of the computing time of the hill climbing method in each iteration. Accuracy in Z-image matching is examined by comparing the matching result with the ground truth, i.e. manually measured location of viewpoints. In Figure 8, estimation error is evaluated by comparing the distance between the relative locations of sensors obtained in Z-image matching with that from the manual measurement data. Estimation error in coarse matching is about 1.43m, while by fine adjustment matching error is about 0.04m (see Figure 8).

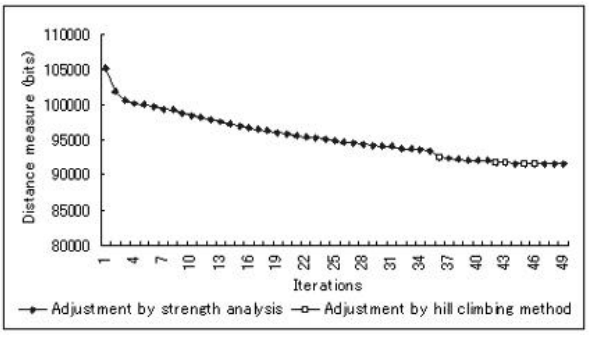

Figure 7. Change of distance measure and comparison of adjustment methods in fine adjustment.

Figure 9 shows the result of matching ground range points to determine the translation parameter along

Z-axis. Gray values of range points on the ground denote height difference between the two Z-images. In Figure 9, it is obvious that the height differences between the ground range points are greatly reduced after the matching of ground range points. Errors of the estimated translation parameter is 0.03m in this case. In Figure 10, the final result of the pair-wise registration of the two range images is shown.

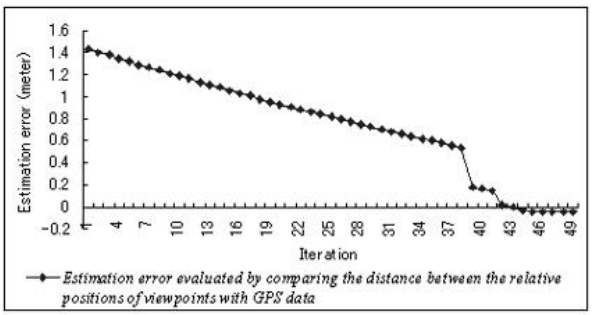

Figure 8. Changes of estimation error in fine adjustment.

Experiment of multiple registration

An experiment was conducted to construct a 3D model of the buildings in the campus of University of Tokyo. 42 overlapping range images were acquired around the major target building. The range images are first sequentially aligned to a global coordinate system based on the results of the pair-wise registration, then simultaneously adjusted to achieve global consistency.

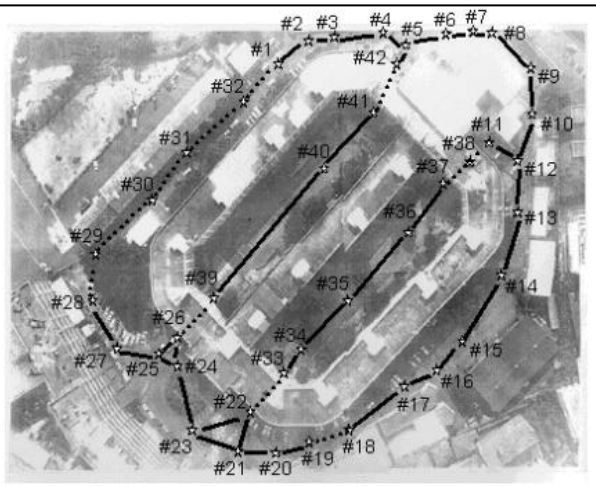

Figure 11. Pair-wise registration of 42 range images of the buildings in the campus of Univ. of Tokyo. *(Star mark denotes the sensor's location of range images, and characters are the sequential number of range images. Solid lines stand for successful pair-wise registration, while dotted lines for failed pairs.)*

Figure 11 shows the sensor's locations of 42 range images by star marks and their sequential numbers. 46 pairs of neighboring range images are first registered to find the inter-frame relations. 36 pairs succeeded (shown in solid lines in Figure 6), while 10 pairs failed (shown in dotted lines in Figure 6). Ratio of success is 78%. Reasons for the failure are 1) lack of linear features for registration; e.g. building surfaces are hidden by large trees in range images (#28~#32,#1); 2) lack of common features between range image pairs due to occlusion or

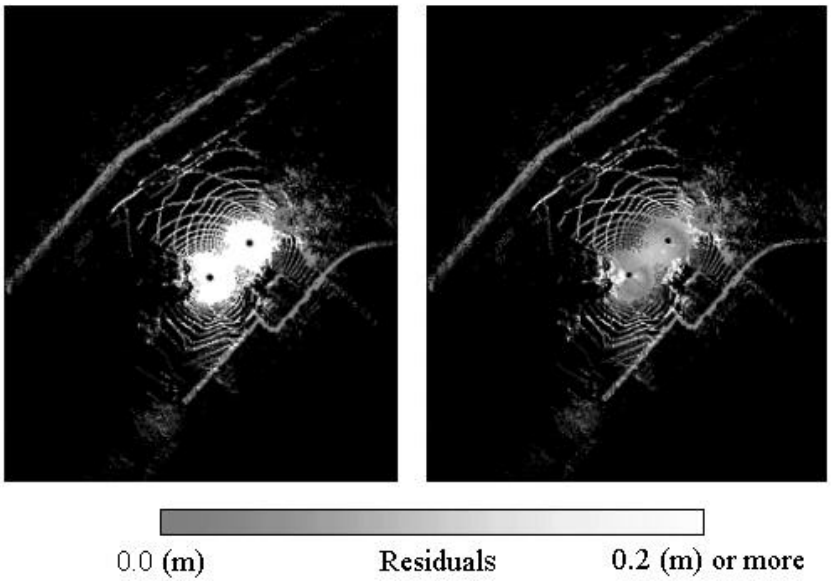

Figure 9. Matching ground range points to determine the translation parameter along Z-axis (a) before matching (b) after matching. *(Matching between tessellated ground surface is shown by intensity values, where perfect matching in the height of ground surface is shown in grey, inaccurate matching which has a residual in the height of ground surfaces equal or larger than 0.2m is shown in white. Circular mark denotes the sensor's location in each range image.)*

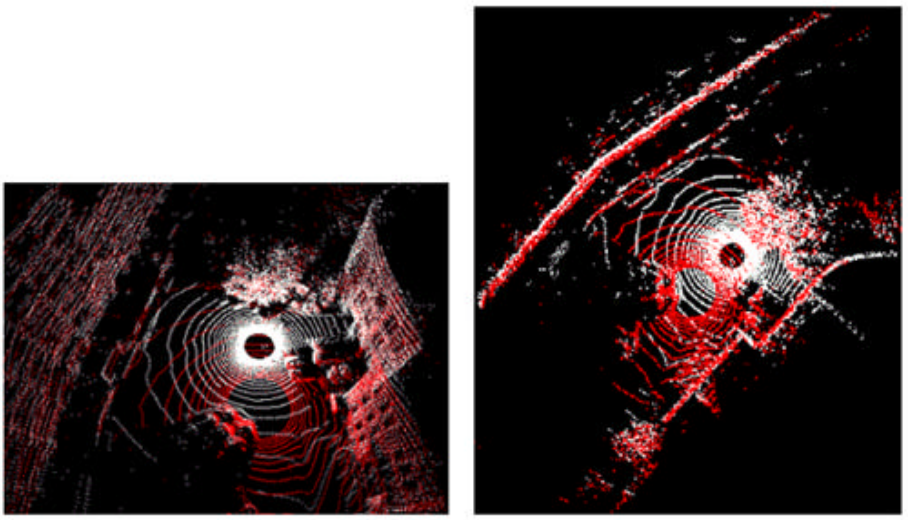

Figure 10. Pair-wise registration result (a) range image (b) Z-Image *(Range image 1 and Z-Image 1 are shown in white, while Range image 2 and Z-Image 2 are shown in red.)*

large distance between neighboring range images, e.g. (#18,#19) and (#28,#29). For those failed in automatic pair-wise registration, we first manually determine an approximate transformation as the initial point, then do automated fine adjustment. It is not difficult to manually determine an approximate transformation when using Z-images, because one can easily slide or rotate a Z-image on one another by keyboard operations. By combining the automated and semi-automated registration methods, all the 42 range images were successfully aligned to the coordinate system of range image #1 (Figure 12(a)). However obvious inconsistencies can be found in the integrated model of range images by sequential alignment.

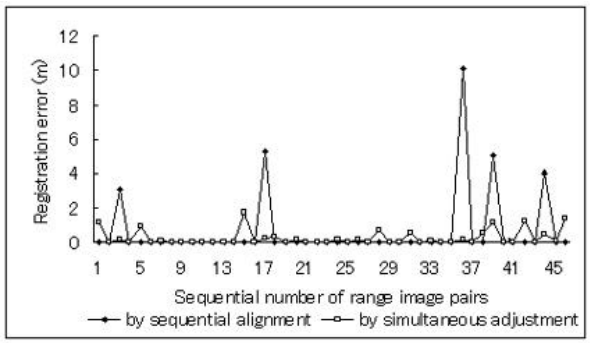

Figure 13. Comparison of the registration errors by sequential alignment and after simultaneous adjustment.

Registration error is evaluated in Figure 13 by $|d_{ij} - \bar{d}_{ij}|$, where d_{ij} and $\bar{d}_{ij}$ (also defined in Figure 5) are the distance between neighboring viewpoints obtained in pair-wise and multiple registration respectively. It can be found that in Figure 13 there are four high peaks after sequential alignment, which represent heavy accumulation of pair-wise registration errors in the four pairs of range images. After simultaneous adjustment, a more consistent model of range images is obtained as shown in Figure 12(b). In Figure 13 it can be found that those accumulated errors are allocated to other pairs of range images after simultaneous adjustment.

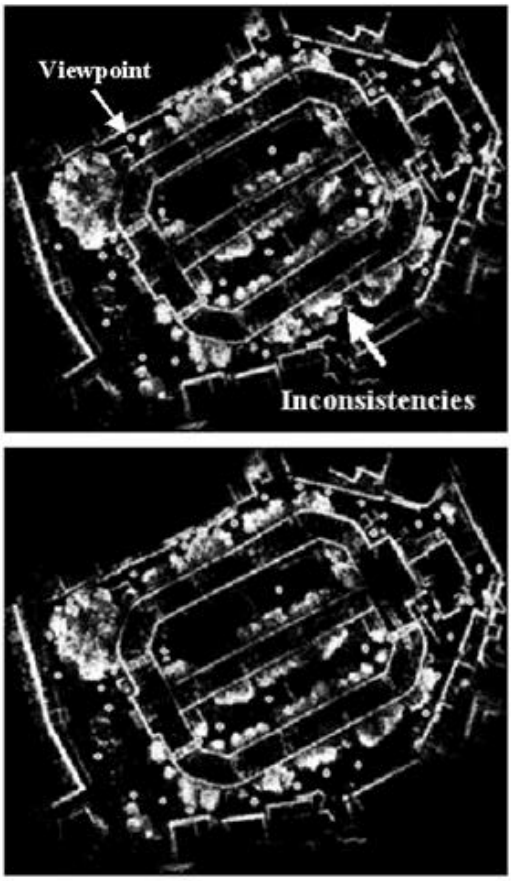

Figure 12. Result of multiple registration (a) sequentially aligning 42 range images using pair-wise registration results (b) result after simultaneous adjustment.

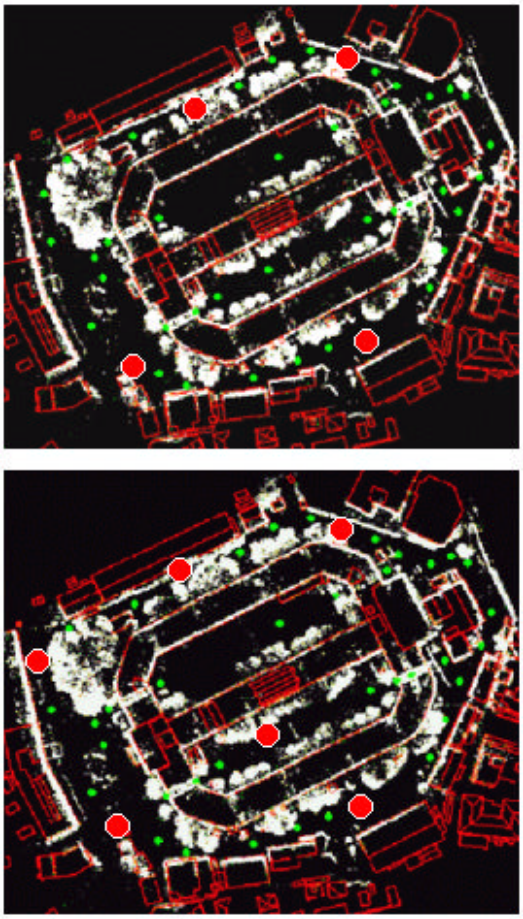

Figure 14. Multiple registration (a) using four GPS points (b) using six GPS points

In order to register all range images directly to a world coordinate system, positions of several viewpoints measured by GPS are exploited. Error of GPS measurement is about ±20cm. Figure 14(a) is the registration result using four GPS measured viewpoints, while Figure 14(b) is the result using six GPS measured viewpoints. Accuracy is examined by comparing the location of 42 viewpoints obtained in multiple registration with those measured by GPS. Residuals between them are shown in Figure 15. Average residual is 0.612m when using four GPS measured viewpoints, while it is 0.283m when using six GPS measured viewpoints. It is seen that the registration error is almost in the same order with GPS measurement when measuring one viewpoint among seven viewpoints using GPS as auxiliary data.

The integrated models of range images are also overlaied with a 1:500 digital map as shown in Figure 14, where a good consistency can be found between them. Residuals from range points to the lines of digital map is examined, and shown in Figure 16. Comparison is made in some selected sites where range points of the building wall can be easily recognized. It can be seen that the residual for a smooth wall is rather small, while that for a wall with projecting edge and window glass is very high. On the other hand it is demonstrated that range images grasp more detailed information than a 1:500 digital map.

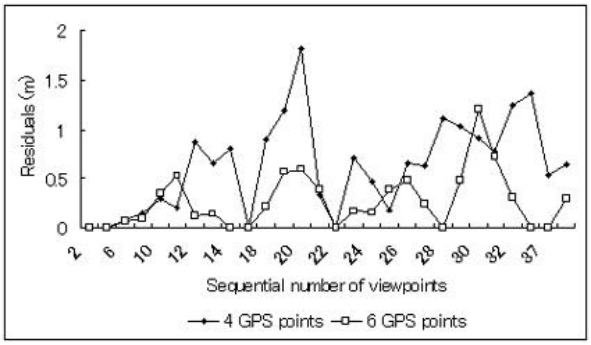

Figure 15. A comparison of registration accuracy when using four and six GPS measured viewpoints

CONCLUSION AND FUTURE STUDY

In this paper, we present a robust method to register multiple overlapping range images of urban objects. The registration is performed at two levels: pair-wise registration and multiple registration. Pair-wise registration is achieved first by horizontal matching of Z-images to determine the horizontal rotation angle and translation parameters in X-Y plane, then by vertical matching of ground range points to determine the translation parameter along Z-axis. Simultaneous registration of multiple range images aims at mitigating the error accumulation in pair-wise registration. Through an outdoor experiment, it is demonstrated that the method has robustness and accuracy in the sense of automation. The paper also contributes to a matching method using *entropy*. Although it was originally motivated for the matching of Z-images, it can be extended to the matching of other types of data sets characterized by geometric features, such as curved lines or planar faces.

Future research is needed on the improvement of accuracy in determining the height of viewpoints, especially in the case when horizontal rotation axis is slanted with respect to the ground. Other topics for future research are to develop a more rigorous method for multiple registration, by integrating the present pair-wise registration method and the multiple registration method, and an automated method to extract popular urban objects such as buildings.

REFERENCE

[1] Bergevin, R., M.Soucy, H.Gagnon, D.Laurendeau,

Towards a General Multi-view Registration Technique, IEEE Trans. PAMI 18(5), pp.540-547, 1996.

[2] Besl, P.J., N.D.McKay, A Method for Registration of 3-D Shapes, IEEE Trans. PAMI vol.14, no.2, pp.239-256, 1992.

[3] Blahut, R.E., Principles and Practice of Information Theory, Addison-Wesley, Reading, MA, 1987.

[4] Boyer, K.L., A.C.Kak, Structural Stereopsis for 3-D Vision, IEEE Trans.PAMI, vol.10, no.2, pp.144-166, 1988.

[5] Champleboux, G., S.Lavallee, R.Szeliski, L.Brunie, From Accurate Range Imaging Sensor Calibration to Accurate Model-Based 3-D Object Localization, Proc. CVPR, 1992, pp.83-88.

[6] Chen, Y., G.Medioni, Object Modelling by Registration of Multiple Range Images, Image and Vision Computing, vol.10, no.3, pp.145-155, 1992.

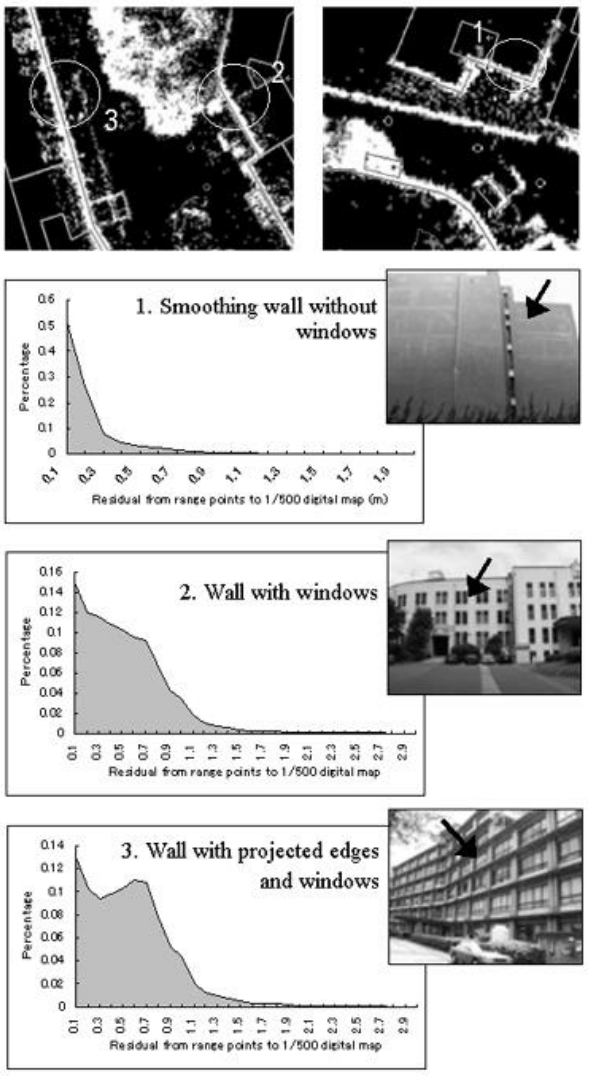

Figure 16. Residuals from range points to 1:500 scale digital map

[7] Forstner, W., Image Analysis Techniques for Digital Photogrammetry, Proc. 42th Photogrammetric Week, Series of the Institute of Photogrammetry, Stuttgart Univ., vol.13, pp.205-221, 1989.

[8] Gruen, A., TOBAGO – a semi-automated approach for the generation of 3-D building models, ISPRS Journal of Photogrammetry and Remote Sensing, vol.53, issue 2, pp.108-118, 1998.

[9] Haala, N., C.Brenner, C.Statter, An integrated system for urban model generation, Int. Archive of Photogrammetry and Remote Sensing, vol.XXXII, part 2, pp.96-103, 1998.

[10] Higuchi, K., M.Hebert, and K.Ikeuchi, Building 3-D Models from Unregistered Range Images, Graphical Models and Image Processing, vol.57, No.4, pp.315-333, 1995.

[11] Ingels, F.M., Information and Coding Theory, In text educational publishers, Scranton, Pennsylvania, 1971.

[12] Jaynes, C., View Alignment of Aerial and Terrestrial Imagery in Urban Environments, ISD'99, LNCS1737, pp.3-19, 1999.

[13] Kamgar-Parsi, J.L.Jones, A.Rosenfeld, Registration of Multiple Overlapping Range Image without Distinctive Features, IEEE Trans.PAMI, vol.13, no.9, pp.857-871, 1991.

[14] Kanatani, K., Geometric Computation for Machine Vision, Oxford University Press, Oxford, U.K., 1993.

[15] Krishnapuram, R., D.Casasent, Determination of Three-Dimensional Object Location and Orientation from Range Images, IEEE Trans.PAMI, vol.11, no.11, pp.1158-1167, 1989.

[16] Lemmens, M.J.P.M., H.Deijkers, P.A.M.Looman, Building detection by fusion airborne laser-altimeter DEMs and 2D digital maps, Int. Archive of Photogrammetry and Remote Sensing, vol.32, Part3-4W2, pp.42-49, 1997.

[17] Liang, P., A new and efficient transform for curve detection, Journal of Robotic Systems, 8(6), pp841-847, 1991.

[18] Ng, K., V.Sequeira, S.Butterfield, D.Hogg, J.G.M.Goncalves, An Integrated Multi-sensory System for Photo-Realistic 3D Scene Reconstruction, Int. Archive of Photogrammetry and Remote Sensing, vol.XXXII, part 5, pp356-363, 1998.

[19] Ozawa, S., M.Notomi, H.Zen, A Wide Scope Modeling to Reconstruct Urban Scene, Int. Archive of Photogrammetry and Remote Sensing, vol.XXXII, part 5, pp370-376, 1998.

[20] Shum, H., K.Ikeuchi, and R.Reddy, Virtual Reality Modeling from a Sequence of Range Images, Proc. IEEE/RSJ Intern. Conf on Intelligent Robots and Systems, pp.703-710, 1994.

[21] Stilla, U., K.Jurkiewicz, Reconstruction of Building Models from Maps and Laser Altimeter Data, ISD'99, LNCS1737, pp.34-46, 1999.

[22] Thorpe,C., M.H.Hebert, T.Kanade, S.A.Shafer, Vision and Navigation for the Carnegie-Mellon Navlab, IEEE Trans. PAMI, vol.10, no.3, pp.363-373, 1988.

[23] Vosselman, G., Relational matching, Lecture notes in computer science 628, Springer-Verlag, Berlin Heidelberg, 1992.

[24] Zhao, H., R.Shibasaki, Automated Registration of Ground-Based Laser Range Image for Reconstructing Urban 3D Object, Int. Archive of Photogrammetry and Remote Sensing, Vol.32, Part3-4W2, pp.27-34,1997.