Robust 2D-3D alignment based on geometrical consistency

Kenji HARA, Yuuki KABASHIMA, Yumi IWASHITA, Ryo KURAZUME, Tsutomu HASEGAWA
Kyushu University, Japan, yumi@irvs.is.kyushu-u.ac.jp

Abstract

This paper presents a new registration algorithm of a 2D image and a 3D geometrical model, which is robust for initial registration errors, for reconstructing a realistic 3D model of indoor scene settings. One of the typical techniques of pose estimation of a 3D model in a 2D image is the method based on the correspondences between 2D photometrical edges and 3D geometrical edges projected on the 2D image. However, for indoor settings, features extracted on the 2D image and jump edges of the geometrical model, which can be extracted robustly, are limited. Therefore, it is difficult to find corresponding edges between the 2D image and the 3D model correctly. For this reason, in most cases, the relative position has to be manually set close to correct position beforehand. To overcome this problem, in the proposed method, firstly the relative pose is roughly estimated by utilizing geometrical consistencies of back-projected 2D photometrical edges on a 3D model. Next, the edge-based method is applied for the precise pose estimation after the above estimation procedure is converged. The performance of the proposed method is successfully demonstrated with some experiments using simulated models of indoor scene settings and actual environments measured by range and image sensors.

1. Introduction

This paper deals with the problem of pose estimation of a 3D geometric model in a 2D image for creating a realistic 3D model of indoor scene settings. Several registration techniques of a 2D image and a 3D geometric model have been proposed so far. One of the typical techniques is the method based on the correspondence of 2D photometrical edges and projected 3D geometrical edges on the 2D image. However, for indoor settings, features extracted on the 2D image and jump edges of the geometrical model, which can be extracted robustly, are limited. Therefore, it is difficult to find corresponding edges between the 2D image and the 3D model correctly. For this reason, in many cases, the relative position has to be manually set close to correct position beforehand, since it is likely to get stuck into local minimums if an initial registration error is large.

In this paper, a new registration algorithm of a 2D image and a 3D geometric model is proposed. The proposed method is able to estimate optimum pose of 3D model in 2D image robustly against initial registration error. In general, most of urban or indoor scene settings consist of multiple planar surfaces, and these planar surfaces and their intersection lines have some geometrical relations with other planar surfaces and intersection lines, respectively. While most of conventional methods proposed so far utilize 2D photometrical edges and projected 3D geometrical edges on the 2D image, the proposed method evaluates geometrical consistencies of back-projecting 2D photometrical edges on the 3D model. By utilizing geometrical properties, such as linearity, parallelism, and orthogonality, the relative pose can be robustly estimated even if an initial registration error is large.

The proposed method consists of two registration algorithms, one is based on geometrical consistency of artificial 3D scene settings [3, 11] and the other is based on correspondences between 2D photometrical and 3D geometrical edges [6]. Firstly, the geometrical consistency-based registration method is applied for determining relative pose robustly against an initial registration error. Next, for the precise pose estimation, the edge-based registration method is applied after the geometrical consistency-based registration method is converged. The performance of the proposed method is successfully demonstrated with some experiments using simulated 3D geometrical models of indoor scene settings and actual environments measured by range and image sensors.

2. Related work

For the precise alignment of 2D images and 3D geometrical models, Kurazume [6] proposed the simultaneous registration algorithm using 2D texture images and reflectance images, which are provided as a side product of range images by most of Time-of-Flight range sensors. In this method, a number of photometrical edges extracted from a texture and a reflectance images are registered and the relative pose between them is determined using robust M-estimator. Epipolar constraints are also utilized to estimate relative poses of multiple texture images simultaneously. Elstrom [4] also used extracted feature points in reflectance images and texture images by a corner detector, and determines correspondence between these feature points. Umeda [16] also utilizes the reflectance image, but this method introduces the optical flow constraint between the reflectance image and the texture image. Intrinsic and extrinsic parameters are determined using the least squares

method.

Other than the methods using reflectance images, some registration techniques using silhouette images or contour lines in 2D image plane have been proposed. Lensch [9], [8], [10] proposed a silhouette-based approach. The size of XOR regions of silhouette image of a 2D image and a projected 3D model is defined as the similarity measure, and the optimum pose which minimizes the size of the XOR regions is determined using the Downhill Simplex method. Brunie [1] and Lavallee [7] utilize a pre-computed 3D distance map of a free form object for 3D pose estimation. The error metric is defined as the minimum distance between the surface of the 3D model and a projection ray, and the sum of the error is minimized using the Levenberg-Marquardt method. To make the registration process efficiently, 3D distance from the surface is pre-computed and stored by the octree structure. Zuffi [17] applies this algorithm for the pose estimation of a knee joint in a single X-ray image for total knee replacement surgery. In the contour-based approach, the error is computed as the sum of distances between points on a contour line in a 2D image and on a projected contour line of a 3D model [2], [13], [14]. Iwashita [5] proposed the fast alignment algorithm utilizing a 2D distance map constructed by Level Set Method. However, features extracted from a 2D image or a 3D model are limited in the indoor settings, and it is likely to get stuck into local minimums if an initial registration error is large.

Liu [12] proposed a registration algorithm aimed at urban scene objects. This method is based on the correspondences of 3D geometrical edges and 2D photometrical edges, but it is likely to get stuck into local minimums in the case that an initial registration error is large. In [15], relative positions between image and range sensors are estimated roughly using vanishing points on 2D images, and the optimum pose is estimated by comparing 2D photometrical edges and 3D geometrical edges projected on the 2D image. However, to estimate vanishing points stably, plenty of parallel lines have to be extracted from the 2D image. Therefore, this method is not suitable for urban environments in which enough number of parallel lines can not be extracted, such as indoor scene settings consisting of a few number of planes.

3. New 2D-3D registration algorithm using geometrical consistencies

In this section, we propose a new registration algorithm of a 2D image and a 3D geometric model which can estimate optimum pose robustly against initial registration error. Hereafter we assume that a 3D geometric model of a scene is constructed and represented by a number of small triangular patches. Intrinsic parameters of an image sensor are also calibrated precisely.

The proposed procedure is summarized as follows:

1. 2D photometrical edges in the 2D image are extracted by feature detectors such as Canny operator.
2. Straight lines and planar surfaces in the 2D image are determined from 2D photometrical edges extracted in step 1.
3. The 3D geometric model is placed at an arbitrary position.
4. Straight lines and planar surfaces are back-projected to the 3D model using the updated parameters of the relative position.
5. Patches of the 3D model on which the straight lines and planar surfaces are back-projected are identified and their 3D positions are calculated by utilizing the high speed rendering function of the OpenGL hardware accelerator.
6. The evaluated value of the geometrical consistencies is calculated and the relative position between the 2D image and the 3D model is estimated using steepest descent method or conjugate gradient method.
7. Step 4 ∼ 6 are repeated until the evaluated value converges.
8. The precise pose is estimated using the edge-based method, which is the method based on the correspondences between 2D photometrical edges and projected 3D geometrical edges on the 2D image.

Step 3 ∼ 7 and Step 8 show the geometrical consistency-based and edge-based registration method, respectively. The above procedure is explained in more details with some examples in the following sections.

3.1. Registration method based on geometrical consistencies

3.1.1 Detection of straight lines and planar surfaces.

Straight lines and planar surfaces in the 2D image are extracted from 2D photometrical edges(Fig.1(a)). Firstly, straight lines are determined with Hough transform or operator's instruction. Next, to determine planar surfaces, a distance map is constructed, which indicates the distance from a pixel in the 2D image to the nearest 2D photometrical edge(Fig.1(b)). By binarizing the distance image with a proper threshold, regions which are far from edges are extracted as planar surfaces(Fig.1(c).

3.1.2 Geometrical consistencies.

In this section, the evaluated value of the geometrical consistencies is explained. The evaluated value indicates the degree of fitness of the geometrical consistencies defined
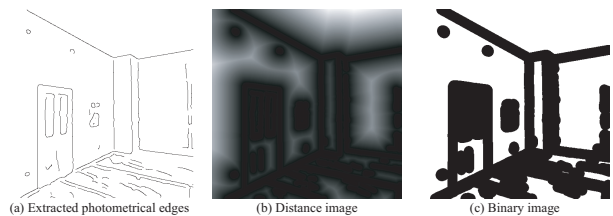

(a) Extracted photometrical edges   (b) Distance image   (c) Binary image

Figure 1. Detection of planar surfaces.

for the back-projected 2D edges on the 3D model. Specifically, this value is defined as the linear sum of the following three evaluated values in this paper.

- Linear consistency
- Planar consistency
- Parallel and orthogonal consistencies

The definitions of these consistencies are explained in more details.

- Linear consistency

  In most cases, a 2D straight line in the 2D image can be regarded as the projection of a 3D straight line in the 3D model. Thus, if the 2D image is aligned to the 3D model precisely, the patches of the 3D model on which the 2D straight line are back-projected should be aligned along a straight line in 3D space. However, if these patches are involved in two different planar surfaces, the back-projected 2D straight line is divided into two straight lines on the 3D model. This is inconsistent with the fact that a 2D straight line corresponds to a single 3D straight line in 3D space (Fig.2). Therefore, by evaluating the degree of the straightness of these 3D patches, the correctness of the 2D-3D registration can be estimated. In this paper, we define this condition as the linear consistency.

  Under the relative position between the 2D image and 3D model $(\mathbf{R}, \mathbf{t})$, the covariance matrix of the 3D positions of the patches of the back-projected straight line SL_i ($0 \leq i < Num_{sl}$, Num_{sl} is the number of detected straight lines.) and eigenvalues of the covariance matrix are calculated. Here, three eigenvalues are defined as $\lambda_{min}(i; \mathbf{R}, \mathbf{t})$, $\lambda_{med}(i; \mathbf{R}, \mathbf{t})$, $\lambda_{max}(i; \mathbf{R}, \mathbf{t})$ ($0 < \lambda_{min}(i; \mathbf{R}, \mathbf{t}) < \lambda_{med}(i; \mathbf{R}, \mathbf{t}) < \lambda_{max}(i; \mathbf{R}, \mathbf{t})$), and eigenvectors are defined as $\mathbf{X}_{min}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{X}_{med}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{X}_{max}(i; \mathbf{R}, \mathbf{t})$, respectively. $\mathbf{X}_{max}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{X}_{med}(i; \mathbf{R}, \mathbf{t})$, and $\mathbf{X}_{min}(i; \mathbf{R}, \mathbf{t})$ indicate the directions corresponding to the maximum, medium, and minimum variances, respectively. Therefore, the smaller the $\lambda_{min}(i; \mathbf{R}, \mathbf{t})$ and $\lambda_{med}(i; \mathbf{R}, \mathbf{t})$ are, the higher the possibility that the back-projected straight line is aligned along a straight line is.

  Thus, the linear consistency is defined as the following equation.

  $$E_{linearity}(\mathbf{R}, \mathbf{t}) = \sum_{0 \leq i < Num_{sl}} \lambda_{min}(i; \mathbf{R}, \mathbf{t}) + \sum_{0 \leq i < Num_{sl}} \lambda_{med}(i; \mathbf{R}, \mathbf{t}) \quad (1)$$

  $\mathbf{X}_{max}(i; \mathbf{R}, \mathbf{t})$ is the directional vector of the approximated 3D line of the back-projected straight line SL_i in 3D space.
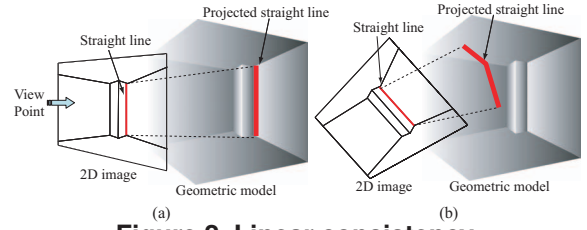

Figure 2. Linear consistency.

- Planar consistency

  In case that the patches on which a planar surface in the 2D image is back-projected are involved in a single planar surface, these patches are aligned evenly on the 3D model. However, if these patches are involved in two different planar surfaces, the back-projected planar surface in the 2D image is divided into two planar surfaces on the 3D model. This is inconsistent with the fact that a 2D planar surface corresponds to a single 3D planar surface in 3D space (Fig.3). Therefore, in the same way as the linear consistency, by evaluating the degree of the smoothness of these 3D patches, the correctness of the 2D-3D registration can be estimated. In this paper, we define this condition as the planar consistency.

  In order to evaluate the planar consistency, the covariance matrix of the 3D positions of the patches of the back-projected planar surface PS_k ($0 \leq k < Num_{ps}$, Num_{ps} is the number of detected planar surfaces.) and eigenvalues of the covariance matrix are calculated. Here, three eigenvalues are defined as $\mu_{min}(k; \mathbf{R}, \mathbf{t})$, $\mu_{med}(k; \mathbf{R}, \mathbf{t})$, $\mu_{max}(k; \mathbf{R}, \mathbf{t})$ ($0 < \mu_{min}(k; \mathbf{R}, \mathbf{t}) < \mu_{med}(k; \mathbf{R}, \mathbf{t}) < \mu_{max}(k; \mathbf{R}, \mathbf{t})$), and eigenvectors are defined as $\mathbf{Y}_{min}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{Y}_{med}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{Y}_{max}(i; \mathbf{R}, \mathbf{t})$, respectively. $\mathbf{Y}_{max}(i; \mathbf{R}, \mathbf{t})$, $\mathbf{Y}_{med}(i; \mathbf{R}, \mathbf{t})$, and $\mathbf{Y}_{min}(i; \mathbf{R}, \mathbf{t})$ indicate the directions corresponding to the maximum, medium, and minimum variances, respectively. Therefore, the smaller the $\mu_{min}(k; \mathbf{R}, \mathbf{t})$ is, the higher the possibility that back-projected planar surface is aligned on a planar surface in the 3D model is.

  Thus, the planar consistency is defined as the following equation.

  $$E_{planarity}(\mathbf{R}, \mathbf{t}) = \sum_{0 \leq k < Num_{ps}} \mu_{min}(k; \mathbf{R}, \mathbf{t}) \quad (2)$$

  $\mathbf{Y}_{min}(i; \mathbf{R}, \mathbf{t})$ is the normal vector of the approximated 3D planar surface of the back-projected planar surface PS_k in 3D space.

- Parallel and orthogonal consistencies

  Parallel and orthogonal consistencies are defined for straight lines and planar surfaces, respectively. In indoor environments, two planar surfaces are usually located in parallel or orthogonal, and thus 2D straight
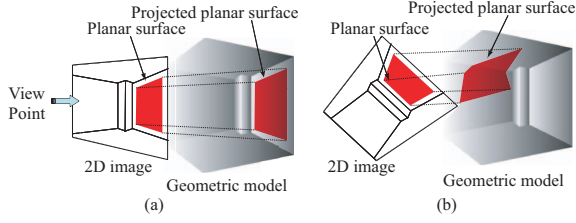
Figure 3. Planar consistency.

lines and 2D planar surfaces have parallel or orthogonal relations with each other in 3D space, respectively. In case that the relative position between image and 3D model is estimated precisely, two back-projected 2D straight lines and two back-projected 2D planar surfaces on the 3D model have parallel or orthogonal relations in most cases. On the other hand, in case that the relative position between them is not estimated correctly, neither parallelism nor orthogonality are satisfied for the back-projected 2D straight lines and 2D planar surfaces on the 3D model (Figs.4 and 5). Therefore, by evaluating the degree of the parallelism or orthogonality of the 3D patches on which the 2D straight lines and 2D planar surfaces are back-projected, the correctness of the 2D-3D registration can be estimated. In this paper, we define these consistencies as the parallel and orthogonal consistencies of straight lines and planar surfaces.

Firstly, the parallel and orthogonal consistencies of straight lines are defined by calculating the eigenvector $\mathbf{X}_{max}(i;\mathbf{R},\mathbf{t})$ and $\mathbf{X}_{max}(j;\mathbf{R},\mathbf{t})$ of the back-projected straight lines SL_i and SL_j $(i > j)$ as the following equations.

$$E_{lineParallelity}(\mathbf{R},\mathbf{t}) \quad = \\ \sum_{i>j} p_{ij} \left\| \mathbf{X}_{max}(i;\mathbf{R},\mathbf{t}) \times \mathbf{X}_{max}(j;\mathbf{R},\mathbf{t}) \right\| \quad (3)$$

$$E_{lineOrthogonality}(\mathbf{R},\mathbf{t}) \quad = \\ \sum_{i>j} q_{ij} \left| \left(\mathbf{X}_{max}(i;\mathbf{R},\mathbf{t}) \cdot \mathbf{X}_{max}(j;\mathbf{R},\mathbf{t}) \right) \right| \quad (4)$$

Here, p_{ij} and q_{ij} $(0 \leq p_{ij}, q_{ij} \leq 1)$ indicate the possibility that the relation between back-projected straight lines SL_i and SL_j has parallel or orthogonal conditions. In case that $p_{ij} = 1$ or $q_{ij} = 1$, two back-projected straight lines have parallel or orthogonal relations, respectively.

Next, the parallel and orthogonal consistencies of planar surfaces are defined by calculating the eigenvector $\mathbf{Y}_{min}(k;\mathbf{R},\mathbf{t})$ and $\mathbf{Y}_{min}(l;\mathbf{R},\mathbf{t})$ of the back-projected planar surfaces PS_k and PS_l $(k > l)$ as the following equations.

$$E_{planeParallelity}(\mathbf{R},\mathbf{t}) \quad = \\ \sum_{k>l} p'_{kl} \left\| \mathbf{Y}_{min}(k;\mathbf{R},\mathbf{t}) \times \mathbf{Y}_{min}(l;\mathbf{R},\mathbf{t}) \right\| \quad (5)$$

$$E_{planeOrthogonality}(\mathbf{R},\mathbf{t}) \quad = \\ \sum_{k>l} q'_{kl} \left| \left(\mathbf{Y}_{min}(k;\mathbf{R},\mathbf{t}) \cdot \mathbf{Y}_{min}(l;\mathbf{R},\mathbf{t}) \right) \right| \quad (6)$$

Here, p'_{kl} and q'_{kl} $(0 \leq p'_{kl}, q'_{kl} \leq 1)$ indicate the possibility that the relation between back-projected planar surfaces PS_k and PS_l has parallel or orthogonal conditions. In case that $p'_{kl} = 1$ or $q'_{kl} = 1$, two back-projected planar surfaces have parallel or orthogonal relations, respectively.
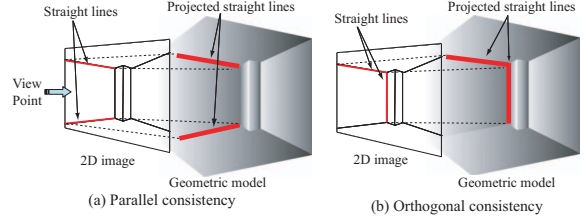

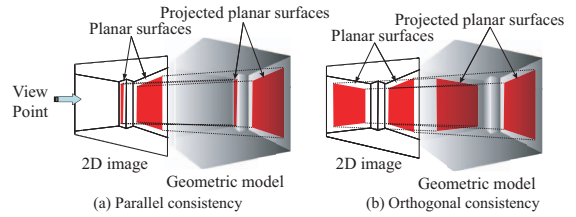**Figure 4. Parallel and orthogonal consistencies of straight lines.**

Figure 5. Parallel and orthogonal consistencies of planar surfaces.

3.1.3 Estimation of relative position and parallel and orthogonal conditions.

The relative position of the 2D image and the 3D model is estimated by minimizing an evaluated value E, which is defined as a linear sum of the evaluated values for the linear, planar, parallel and orthogonal consistencies shown in Eqs.(1) $\sim$ (6).

$$E(\mathbf{R},\mathbf{t}) \quad = \quad E_{linearity}(\mathbf{R},\mathbf{t}) + \alpha_1 E_{lineParallelity}(\mathbf{R},\mathbf{t}) \\ + \alpha_2 E_{lineOrthogonality}(\mathbf{R},\mathbf{t}) + \alpha_3 E_{planarity}(\mathbf{R},\mathbf{t}) \\ + \alpha_4 E_{planeParallelity}(\mathbf{R},\mathbf{t}) + \alpha_5 E_{planeOrthogonality}(\mathbf{R},\mathbf{t}) \\ (7)$$

In addition to the above procedure, it is possible to estimate the parallel and orthogonal conditions at the same time by the following procedure. Firstly, $\mathbf{P} = \{p_{ij}\}_{i>j}$, $\mathbf{Q} = \{q_{ij}\}_{i>j}$, $\mathbf{P}' = \{p'_{kl}\}_{k>l}$, and $\mathbf{Q}' = \{q'_{kl}\}_{k>l}$ are regarded as unknowns, and Eqs.(3) $\sim$ (6) are redefined as the following equations.

$$E_{lineParallelity}(\mathbf{R},\mathbf{t},\mathbf{P}) \quad = \\ \sum_{i>j} p_{ij} \left\| \mathbf{X}_{max}(i;\mathbf{R},\mathbf{t}) \times \mathbf{X}_{max}(j;\mathbf{R},\mathbf{t}) \right\| + c_1 \sum_{i>j} (1.0 - p_{ij})$$

$$E_{lineOrthogonality}(\mathbf{R},\mathbf{t},\mathbf{Q}) \quad =$$

$$\sum_{i>j} q_{ij} \left| (\mathbf{X}_{max}(i\,;\mathbf{R},\mathbf{t}) \cdot \mathbf{X}_{max}(j\,;\mathbf{R},\mathbf{t})) \right| + c_2 \sum_{i>j}(1.0 - q_{ij})$$

$$E_{planeParallelity}(\mathbf{R},\mathbf{t},\mathbf{P}') =$$
$$\sum_{k>l} p'_{kl} \left\| \mathbf{Y}_{min}(k\,;\mathbf{R},\mathbf{t}) \times \mathbf{Y}_{min}(l\,;\mathbf{R},\mathbf{t}) \right\| + c_3 \sum_{k>l}(1.0 - p'_{kl})$$

$$E_{planeOrthogonality}(\mathbf{R},\mathbf{t},\mathbf{Q}') =$$
$$\sum_{k>l} q'_{kl} \left| (\mathbf{Y}_{min}(k\,;\mathbf{R},\mathbf{t}) \cdot \mathbf{Y}_{min}(l\,;\mathbf{R},\mathbf{t})) \right| + c_4 \sum_{k>l}(1.0 - q'_{kl})$$
$$(8)$$

where $c_1 \sim c_4$ are constants. By adding the new term to Eqs.(3) $\sim$ (6) and minimizing Eq.8, p_{ij}, q_{ij}, p'_{kl}, and q'_{kl} converge to 0 or 1 after iterative estimation. Therefore the evaluated value E is defined as follows:

$$E(\mathbf{R},\mathbf{t},\mathbf{P},\mathbf{Q},\mathbf{P}',\mathbf{Q}') = E_{linearity}(\mathbf{R},\mathbf{t})$$
$$+\alpha_1 E_{lineParallelity}(\mathbf{R},\mathbf{t},\mathbf{P}) + \alpha_2 E_{lineOrthogonality}(\mathbf{R},\mathbf{t},\mathbf{Q})$$
$$+\alpha_3 E_{planarity}(\mathbf{R},\mathbf{t}) + \alpha_4 E_{planeParallelity}(\mathbf{R},\mathbf{t},\mathbf{P}')$$
$$+\alpha_5 E_{planeOrthogonality}(\mathbf{R},\mathbf{t},\mathbf{Q}') + \alpha_6 E_{linePenalty}(\mathbf{P},\mathbf{Q})$$
$$+\alpha_7 E_{planePenalty}(\mathbf{P}',\mathbf{Q}') \qquad (9)$$

where $\alpha_1 \sim \alpha_7$ are constants, and $E_{linePenalty}$ and $E_{planePenalty}$ are additional constraints that two straight lines or two planar surfaces have never parallel and orthogonal relations with each other at the same time. These terms are defined as $E_{linePenalty}(\mathbf{P},\mathbf{Q}) = \sum_{i>j} p_{ij} q_{ij}$ and $E_{planePenalty}(\mathbf{P}',\mathbf{Q}') = \sum_{k>l} p'_{kl} q'_{kl}$, respectively. At the beginning of the estimation process, parallel and orthogonal conditions are set to $p_{ij} = q_{ij} = p'_{kl} = q'_{kl} = 0.5$ as the initial values, and are estimated using intermediate parameters $m_{ij}, n_{ij}, m'_{kl}, n'_{kl}$ as follows:

$$m_{ij,t} = m_{ij,t-1} - \frac{\partial E}{\partial p_{ij}} \qquad n_{ij,t} = n_{ij,t-1} - \frac{\partial E}{\partial q_{ij}}$$

$$m'_{kl,t} = m'_{kl,t-1} - \frac{\partial E}{\partial p'_{kl}} \qquad n'_{kl,t} = n'_{kl,t-1} - \frac{\partial E}{\partial q'_{kl}} \quad (10)$$

where t is the iteration number of convergent calculation, and p_{ij}, q_{ij}, p'_{ij}, and q'_{ij} are defined by the sigmoid function.

$$p_{ij} = \frac{1}{1 + e^{-\frac{m_{ij}}{S_T}}} \qquad q_{ij} = \frac{1}{1 + e^{-\frac{n_{ij}}{S_T}}}$$

$$p'_{kl} = \frac{1}{1 + e^{-\frac{m'_{kl}}{S_T}}} \qquad q'_{kl} = \frac{1}{1 + e^{-\frac{n'_{kl}}{S_T}}} \qquad (11)$$

where S_T is a constant of the sigmoid function. By using the sigmoid function, the range of p_{ij}, q_{ij}, p'_{kl} and q'_{kl} can be limited from 0 to 1. In our implementation, we minimize the evaluated value E in Eq.(9) by steepest descent method or conjugate gradient method.

3.2. Precise registration based on edge correspondences

After the geometrical consistency-based registration is converged, the edge-based registration is applied from the estimated relative position. The edge-based registration is based on correspondences between 2D photometrical and 3D geometrical edges [6]. In this method, photometrical and geometrical edges extracted from a texture image and a 3D geometric model are registered and the relative position is estimated so that these edges coincide with each other. In our implementation, we use straight lines defined in section 3.1 as photometrical edges and geometrical edges are extracted by detecting intersection lines between planar surfaces.

First, geometrical edges are projected to the texture image as shown in Fig.6, and nearest photometrical edges from geometrical edges are searched by using the k-D tree structure. Then, the evaluated value E_{de} is defined as the following equation using the distance d_i[pixel] between these edge points. We define this value as the distance error.

$$E_{de}(\mathbf{R},\mathbf{t}) = \frac{\sum_i^{Num_{de}} d_i}{Num_{de}} \qquad (12)$$

where Num_{de} is the number of the projected geometrical edge points to the texture image.

To update parameters of the relative position between the 2D image and the 3D model, a force $\mathbf{f_i}$ is applied to the geometrical edge point $\mathbf{P}$. The force $\mathbf{f_i}$ is the vector which is on a perpendicular line from the geometrical edge point $\mathbf{P}$ to the stretched line between the optical center and the photometrical edge point. Using the force $\mathbf{f_i}$, current pose $\mathbf{R}$ and position $\mathbf{t}$ are updated as follows:

$$\mathbf{R} \leftarrow \mathbf{R} + \alpha \sum_i^{Num_{de}} \mathbf{f}_i \times (\mathbf{P} - \mathbf{G})$$

$$\mathbf{t} \leftarrow \mathbf{t} + \beta \sum_i^{Num_{de}} \mathbf{f}_i \qquad (13)$$

where α and β are constants, and $\mathbf{G}$ is the center of gravity of the 3D model. These procedures are repeated until the distance error E_{de} converges.
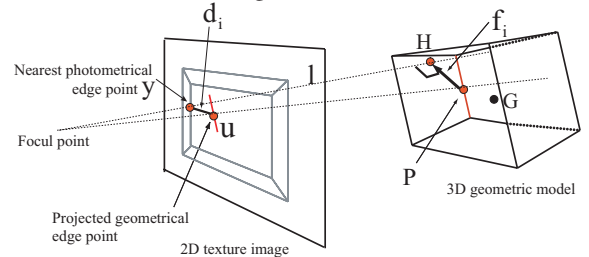

Figure 6. Distance error between the photometrical and geometrical edges.

4. Experiments

In this section, we show some experimental results of the 2D-3D registration using a simulated 3D model and actual environments measured by range and image sensors. In the simulation, we focus on the convergence performance of

the proposed method against local minimums and do not consider the effect of noise on observations.

4.1. Comparison of the accuracy of geometrical consistency-based and edge-based registration methods

Firstly, we examined the change of the evaluated values of the geometrical consistencies(Eq.(7)) and the distance error(Eq.(12)) for various relative positions between image and 3D models. Here, the parallel and orthogonal conditions of straight lines and planar surfaces are assumed to be known. The simulated 3D model shown in Fig.7 (a) is an indoor environment which is 5.0 meters high, 15.0 meters width, and 10.0 meter depth, and Fig.7 (b) shows the 2D image obtained by capturing the textured 3D model. In these experiments, the pose and position errors between the 2D image and the 3D model are set to $(\theta_{roll}, \theta_{pitch}, \theta_{yaw})$[degree] and (t_x, t_y, t_z)[m], where $-30 \le \theta \le 30$ and $-3 \le t \le 3$.

Figure 8(a) shows the evaluated values and Fig.8(b) is a close up of Fig.8(a) around the correct position. The horizontal axis indicates the pose error $(\pm\sqrt{\theta_{roll}^2 + \theta_{pitch}^2 + \theta_{yaw}^2})$ and position error $(\pm\sqrt{t_x^2 + t_y^2 + t_z^2})$, and the left and right vertical axes indicate the evaluated value E of the geometrical consistencies and the distance error E_{de}, respectively. From Fig.8(a), though the distance error E_{de} has many local minimums, we can see the evaluated value E is decreasing almost monotonically to the correct position. This means that in the case that an initial registration error is large, the relative position between the 2D image and the 3D model can be estimated more robustly by the geometrical consistency-based registration method comparing with the edge-based registration method. On the other hand, as shown in Fig.8(b), the distance error converges steeply around the correct position, while the evaluated values E converges gradually. This indicates the edge-based registration method is more suitable for the precise alignment than the geometrical consistency-based registration method. From these consideration, we can conclude that the proposed method which combines these two methods has both properties of robustness for initial registration error and high accuracy for precise registration.
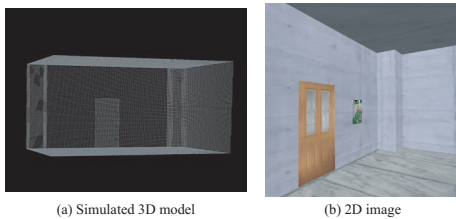

(a) Simulated 3D model   (b) 2D image

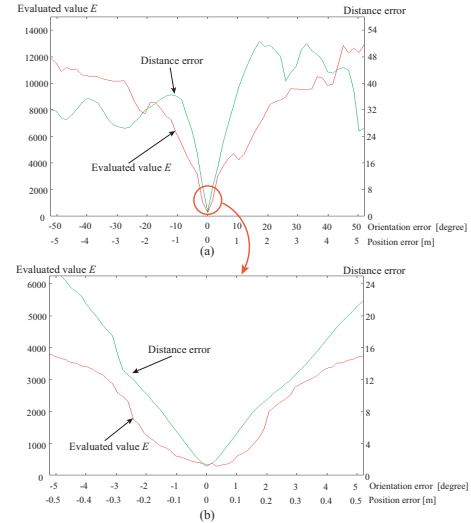**Figure 7. 3D geometrical model and its textured image.**

Figure 8. Comparison of evaluated value of Eqs.7 and 12.

4.2. Evaluation of robustness against initial registration errors

4.2.1 2D-3D registration in case that parallel and orthogonal conditions are known.

Next, we conduct the 2D-3D registration experiments using the simulated 3D model on the condition that the parallel and orthogonal relations of straight lines and planar surfaces are known. The performance of the convergence of three methods, that is, the geometrical consistency-based registration method, the edge-based registration method, and the proposed method, is compared. In these experiments, the initial position is changed 100 times in a random manner. The maximum values of the initial registration errors are 2 meters and 20 degrees for each axis. Figure 9 shows the distribution of the frequencies of position and orientation errors after convergent calculation by the geometrical consistency-based registration method, the edge-based registration method, and the proposed method. Examples of the experimental results with the edge-based registration method and the proposed method are shown in Figs.10 and 11.

As seen in these results, the edge-based registration method estimates the relative position between the 2D image and the 3D model precisely if the relative position is converged to the correct position. However, in the case that the initial position is far from the correct position, it is likely to get stuck into local minimums. On the other hand, though the accuracy of the geometrical consistency-based registration algorithm is lower than the edge-based registration method, the possibility of local minimums of the proposed method is lower than the edge-based registration method. In the proposed method, the geometrical-based method could estimate the relative position close to the correct position firstly, and then the edge-based method

was applied for the precise pose estimation from the initial position estimated by the geometrical-based method. Hence the proposed method estimated the relative position precisely, and the success rate in which the distance error becomes under 1.5 [pixel] is 96.0 %.
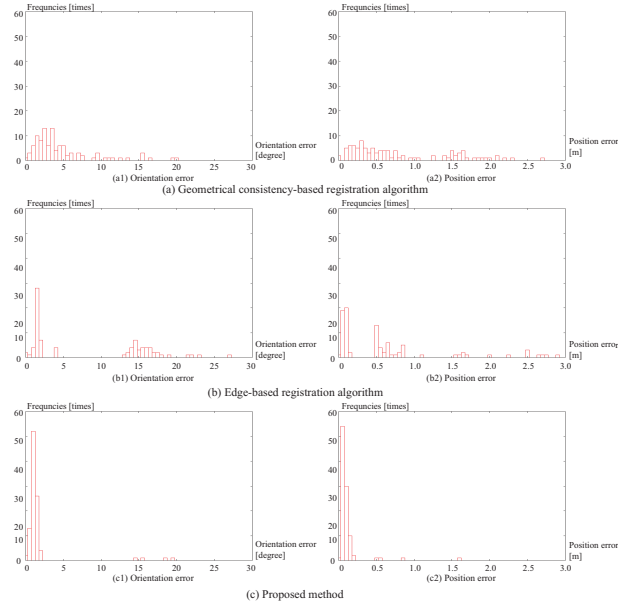

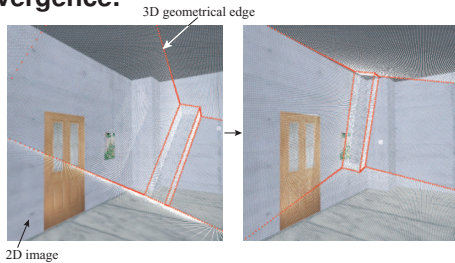**Figure 9. Distribution of the frequencies after convergence.**

Figure 10. 2D-3D registration with the edge-based registration method only.

4.2.2 2D-3D registration in case that parallel and orthogonal conditions are also estimated.

We conduct the experiments to estimate the parallel or orthogonal conditions of straight lines and planar surfaces automatically by minimizing the evaluated value E detailed in Eq.(9).

Figure 12 shows the distribution of the frequencies of position and orientation errors after convergent calculation by the proposed method. The success rate in which the distance error becomes under 1.5 [pixel] is 75.0 %. Figure 13 shows examples of the estimated results of the parallel and orthogonal conditions. These results show that the parallel and orthogonal conditions are estimated correctly. For all of the straight lines, the success rates of the estimation of the parallel and orthogonal conditions are 85.7 % and 83.5 %, respectively. For all of the planar surfaces, the success rates are 100.0 %.
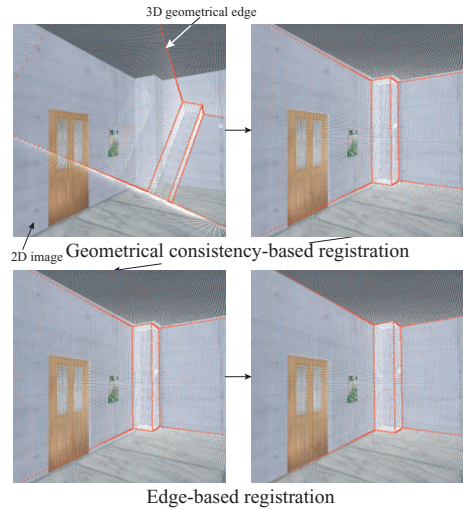

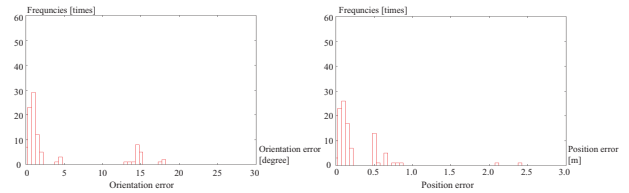**Figure 11. 2D-3D registration with the proposed method.**

Figure 12. Distribution of the frequencies after convergence.

4.3. 2D-3D registration using actual models

Finally, we carry out an experiment of 2D-3D registration using an actual environment measured by a range sensor (SICK, LMS200) and an image sensor (FUJIFILM, FinePix). In this experiment, the condition of parallel and orthogonal relations are assumed to be unknown and estimated. Figures 14 (a) and 14 (b) show the actual 3D geometrical model of a room and its corresponding 2D image. Figure 14 (c) shows the experimental results of estimation of relative pose between the 2D image and the 3D model. From these experiments, we can conclude that that 2D-3D registration is performed successfully for actual environments.

5. Conclusion

This paper proposed a new registration algorithm of a 2D image and a 3D geometrical model for creating realistic 3D model of indoor scene settings. In this method, firstly the relative position between the 2D image and the 3D model is estimated utilizing geometrical consistencies of back-projected 2D photometrical edges on the 3D model. Next, the edge-based method is applied after the geometrical consistency-based registration method is converged for the optimum pose estimation.

Our registration algorithm realizes the robust estimation of relative position between the 2D image and the 3D model against initial registration errors. This method works well even if the features extracted from a 2D image or a 3D
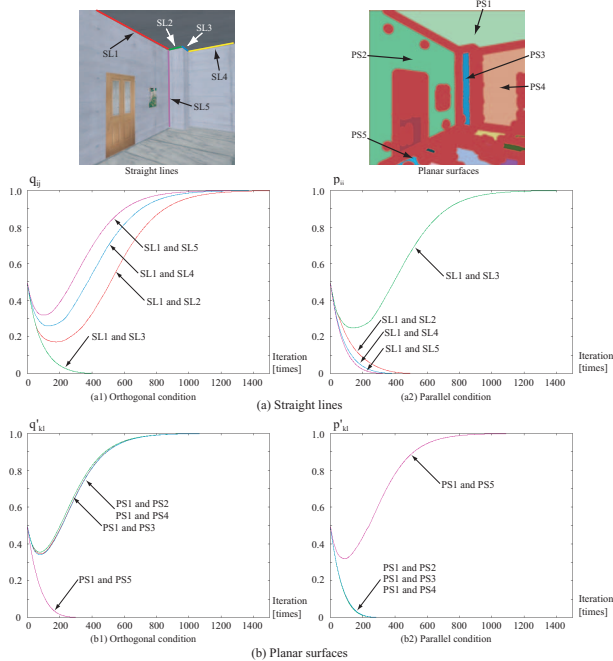
Figure 13. Estimation of parallel and orthogonal conditions of straight lines and planar surfaces.

Figure 14. 2D-3D registration using actual models.

model are limited such as for indoor environments. The relevance of the proposed method was verified through the experiments using simulated 3D models of indoor scene settings and actual environments measured by image and range sensors.

References

[1] L. Brunie, S. Lavallee, and R. Szeliski. Using force fields derived from 3d distance maps for inferring the attitude of a 3d rigid object. *Proc. of the Second European Conference on Computer Vision*, pages 670–675, 1992.

[2] Q. Delamarre and O. Faugeras. 3d articulated models and multi-view tracking with silhouettes. *Proc. of the Int. Conf. Computer Vision*, 2:716–721, 1999.

[3] P. E. Devebec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs:a hybrid geometry- and image-based approach. *Proc. of ACM SIGGRAPH'96*, pages 11–20, 1996.

[4] M. D. Elstrom and P. W. Smith. Stereo-based registration of multi-sensor imagery for enhanced visualization of remote environments. *Proc. of the 1999 IEEE Int. Conf. Robotics and Automation*, pages 1948–1953, 1999.

[5] Y. Iwashita, R. Kurazume, K. Hara, and T. Hasegawa. Fast alignment of 3d geometrical models and 2d color images using 2d distance maps. *Proc. of The 5th Int. Conf. 3-D Digital Imaging and Modeling*, pages 164–171, 2005.

[6] R. Kurazume, K. Nishino, Z. Zhang, and K. Ikeuchi. Simultaneous 2d images and 3d geometric model registration for texture mapping utilizing reflectance attribute. *Proc. of Fifth Asian Conference on Computer Vision*, pages 99–106, 2002.

[7] S. Lavallee and R. Szeliski. Recovering the position and orientation of free-form objects from image contours using

3d distance maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):378–390, 1995.

[8] H. Lensch, W. Heidrich, and H. P. Seide. Hardware-accelerated silhouette matching. *In SIGGRAPH Sketches*, 2000.

[9] H. Lensch, W. Heidrich, and H.-P. Seidel. Automated texture registration and stitching for real world models. *In Pacific Graphics '00*, pages 317–326, 2000.

[10] H. Lensch, W. Heidrich, and H.-P. Seidel. A silhouette-based algorithm for texture registration and stitching. *Graphical Models*, 63:245–262, 2001.

[11] H. Lipson and M. Shpitalni. Optimization-based reconstruction of a 3d object from a single freehand line drawing. *J. Computer Aided Design*, 28(8):651–663, 1996.

[12] L. Liu and I. Stamos. Automatic 3d to 2d registration for the photorealistic rendering of urban scenes. *Proc. of Computer Vision and Pattern Recognition*, 2:137–143, 2005.

[13] K. Matsushita and T. Kaneko. Efficient and handy texture mapping on 3d surfaces. *Comput. Graphics Forum 18*, pages 349–358, 1999.

[14] P. Neugebauer and K. Klein. Texturing 3d models of real world objects from multiple unregistered photographic views. *Comput. Graphics Forum 18*, pages 245–256, 1999.

[15] I. Stamos and P. K. Allen. Automatic registration of 2-d with 3-d imagery in urban environments. *Proc. of the Int. Conf. Computer Vision*, pages 731–737, 2001.

[16] K. Umeda, G. Godin, and M. Rioux. Registration of range and color images using gradient constraints and range intensity images. *Proc. of 17th Int. Conf. Pattern Recognition*, pages 12–15, 2004.

[17] S. Zuffi, A. Leardini, F. Catani, S. Fantozzi, and A. Cappello. A model-based method for the reconstruction of total knee replacement kinematics. *IEEE Trans. on Medical Imaging*, 18(10):981–991, 1999.