# Registration of 3D Geometric Model and Color Images Using SIFT and Range Intensity Images

Ryo Inomata[1], Kenji Terabayashi[1], Kazunori Umeda[1], and Guy Godin[2]

[1] Chuo University,
1–13–27 Kasuga, Bunkyo-ku, Tokyo 112–8551, Japan
inomata@sensor.mech.chuo-u.ac.jp, {terabayashi,umeda}@mech.chuo-u.ac.jp
[2] National Research Council,
1200 Montreal Road, Ottawa, Ontario K1A 0R6, Canada

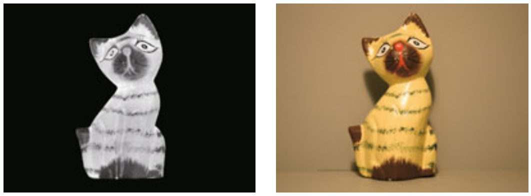
**Abstract.** In this paper, we propose a new method for 3D-2D registration based on SIFT and a range intensity image, which is a kind of intensity image simultaneously acquired with a range image using an active range sensor. A linear equation for the registration parameters is formulated, which is combined with displacement estimations for extrinsic and intrinsic parameters and the distortion of a camera's lens. This equation is solved to match a range intensity image and a color image using SIFT. The range intensity and color images differ, and the pairs of matched feature points usually contain a number of false matches. To reduce false matches, a range intensity image is combined with the background image of a color image. Then, a range intensity image is corrected for extracting good candidates. Moreover, to remove false matches while keeping correct matches, soft matching, in which false matches are weakly removed, is used. First, false matches are removed by using scale information from SIFT. Secondly, matching reliability is defined from the Bhattacharyya distance of the pair of matched feature points. Then RANSAC is applied. In this stage, its threshold is kept high. In our approach, the accuracy of registration is advanced. The effectiveness of the proposed method is illustrated by experiments with real-world objects.

## 1 Introduction

As information technology is becoming more widely used, research that creates realistic models using computer graphics techniques is increasing in importance [1]. Texture mapping on scanned real world objects, which is the method used to map texture images measured using a color sensor on 3D geometric model measured by a range sensor, is one of the methods for creating realistic models. Usually, a 3D geometric model and color image are independently obtained from different viewing positions through range and color sensors. Thus, the registration of a 3D geometric model and color images is necessary.

One approach is the method of registration using a silhouette image and a 2D image contour [2], [3], [4], [5]. The concept is that a silhouette image from 3D geometry is compared with a 2D image. Iwakiri et al. [2] proposed a fast

G. Bebis et al. (Eds.): ISVC 2011, Part I, LNCS 6938, pp. 325–336, 2011.



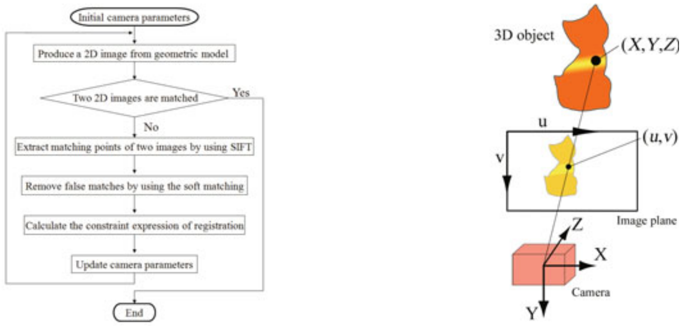
(a) Range intensity image (b) Color image

**Fig. 1.** Range intensity image and color image

method using hierarchical silhouette matching. Neugebauer et al. [5] estimated camera parameters by matching the features of a 3D model and a 2D image by hand. On the other hand, one category of registration techniques relies on the use of range intensity images. An active range sensor measures distance by projecting light and measuring the property of the reflected light. The amount of reflected light is related to the reflectance ratio of the measured points. This is called "reflectance image" or "range intensity image." The range intensity image is similar to a color image as shown in Fig. 1. These images are used in the 3D-2D registration. The similarity of two images is used by the following approaches. Boughorbel et al. [6] used the $\chi^2$-similarity metric to measure the similarity between the range intensity image and the intensity image. Umeda et al. [7] proposed a method for estimating the registration of a range sensor and a color sensor on the basis of the gradient constraint between a range intensity image and a 2D image.

Other approaches use the following features: edges (Kurazume et al. [8]), corners (Elstrom et al. [9]), and SIFT (Scale-Invariant Feature Transform) [11]. SIFT has been shown to be largely invariant to scale changes and varying illumination; thus, the feature is a good candidate for registration. Bohm et al. [10] match a range intensity image and a color image using SIFT. Then, the rigid body transform for a pair-wise registration is computed. The problem is that only extrinsic parameters are estimated.

In our approach, using SIFT, the intrinsic and extrinsic parameters and the distortion of the camera's lens are estimated for highly accurate registration. The range intensity and color images differ, and the pair of matched feature points usually contains a number of false matches. For highly accurate registration, we propose a method to reduce false matches, i.e., soft matching. In this paper, soft matching means that false matches are weakly removed while correct matches are kept. To reduce false matches, a range intensity image is made similar to a color image before using SIFT. First, a range intensity image is combined with a background image of a color image. Then, a range intensity image is corrected for extracting good candidates. Moreover, soft matching is used. Soft matching consists of the following three methods: matching reliability from the Bhattacharyya distance of matched feature points, using scale information from SIFT, and RANSAC [12] is applied. In this stage, its threshold is kept high. As a result, the accuracy of registration is advanced. In this paper, false matches are reduced, and




(a) Flow of the 3D-2D registration (b) Projection of a 3D point on an image plane

**Fig. 2.** Flow of the 3D-2D registration and projection of a 3D point on an image plane

two images are matched using SIFT. Then, false matches are removed, and a linear equation as parameters for the 3D-2D registration is solved. Finally, the camera parameters and the distortion of the camera's lens are updated with the obtained correction of the parameters.

This paper is organized as follows. In Section 2, we show the flow of the 3D-2D registration. In Section 3, a linear equation for the 3D-2D registration is formulated. After introducing the method to reduce false matches in Section 4, we show the soft matching in Section 5. We show several experimental results in Section 6, and we conclude the paper in Section 7.

## 2 Outline of the Registration Method

The inputs of the 3D-2D registration are a color image and a 3D geometric model with range intensity images. For the registration of the 3D geometric model and a color image, it is necessary to obtain the camera parameters, i.e., the intrinsic and the extrinsic ones, in the coordinate system of the 3D geometric model. If the camera parameters are appropriate, the color image and the range intensity image projected on the camera's image plane are matched. In other words, 3D-2D registration entails obtaining the camera parameters to match two images. In practice, the distortion of the camera's lens should also be considered for high-accuracy registration. Fig. 2 (a) shows the flow of the registration. First, the initial values of the parameters are given. A range intensity image is projected on the camera's image plane using the camera parameters, producing a 2D image. The projection is applied to the 3D coordinates of the range image points. The projected range intensity image is compared with the color image. If they are not sufficiently matched, a range intensity image is made similar to a color image. After using SIFT, soft matching is used. The camera parameters are then updated with the obtained correction of the parameters. We iteratively apply the loop process in three stages. In stage 1, only extrinsic parameters are updated. In stage 2, the translation velocity vector and intrinsic parameters are updated. In stage 3, the camera parameters and the distortion of the camera's lens are updated. To evaluate the matching of the range intensity image and the color image, we choose a correlation coefficient.

In order to obtain a color image which is similar to the range intensity image, the one color channel, similar to the projecting light of a range sensor, of the color image is used for comparison with the range intensity image.

## 3 Formulation of the 3D-2D Registration

A 3D point $(X, Y, Z)$ is projected at $(u, v)$ on the image plane of the camera (see Fig. 2 (b));

$$u = \frac{\alpha_u X + sY}{Z} + u_0, \qquad v = \frac{\alpha_v Y}{Z} + v_0, \tag{1}$$

where $\alpha_u, \alpha_v, u_0, v_0, s$ are the intrinsic parameters. $\alpha_u, \alpha_v$ are the aspect ratio, $u_0, v_0$ are the center of the image, and $s$ is the skew.

Firstly, we formulate a linear equation for extrinsic parameters. Then, we extend the equation to the case of variable intrinsic parameters and distortion of the camera's lens.

### 3.1 Constraint for Extrinsic Parameters

Suppose that the intrinsic parameters are given and constant. In this case, Eq. (1) is differentiated to

$$\dot{u} = \frac{\alpha_u}{Z}\dot{X} + \frac{s}{Z}\dot{Y} - \frac{\alpha_u X + sY}{Z^2}\dot{Z}, \tag{2}$$

$$\dot{v} = \frac{\alpha_v}{Z}\dot{Y} - \frac{\alpha_v Y}{Z^2}\dot{Z}. \tag{3}$$

In this paper, the time dimension for derivatives refers to the difference between the projected range intensity image and the color image, which has a color component near the projected light of a range sensor; a color image is virtually translated from a range intensity image in the unit time. Therefore, by using SIFT, $\dot{u}, \dot{v}$ is a translation of a color image from a range intensity image.

Eq. (4) is given by $\dot{\mathbf{X}} = [\dot{X}, \dot{Y}, \dot{Z}]^T$; i.e., the velocity of the 3D point results from the camera motion.

$$\dot{\mathbf{X}} = -\mathbf{v} - \omega \times \mathbf{X}, \tag{4}$$

where $\mathbf{v} = [v_x, v_y, v_z]^T$ is the translation velocity vector and $\omega = [\omega_x, \omega_y, \omega_z]^T$ is the rotation velocity vector. In this paper, these parameters are referred to as motion parameters. By substituting Eq. (4) in Eq. (2) and Eq. (3),

$$\dot{u} = -av_x - bv_y - cv_z - (cY - bZ)\omega_{\mathbf{x}}$$

$$-(aZ - cX)\omega_{\mathbf{y}} - (bX - aY)\omega_{\mathbf{z}}, \tag{5}$$

$$\dot{v} = -dv_y - ev_z - (eY - dZ)\omega_{\mathbf{x}}$$

$$+eX\omega_{\mathbf{y}} - dX\omega_{\mathbf{z}}, \tag{6}$$

where $a, b, c, d$, and $e$ are given by

$$a = \frac{\alpha_u}{Z}, b = \frac{s}{Z}, c = -\frac{\alpha_u X + sY}{Z^2}, d = \frac{\alpha_v}{Z}, e = -\frac{\alpha_v Y}{Z^2}.$$

These are linear equations for the six motion parameters $\mathbf{v}$ and $\omega$. Therefore, they can be calculated by obtaining the three or more points and applying the simple linear least squares method.

In practice, small displacements are used for velocity components in Eq. (4). Extrinsic parameters are represented by $3 \times 3$ rotation matrix $R$ and 3D translation vector $\mathbf{t} = [t_x, t_y, t_z]^T$. They are directly calculated from the obtained small displacements. Eq. (4) is for small displacements, the correct values of the extrinsic parameters are usually not reached on the first try. Therefore, the loop process is iteratively applied (see Fig. 2 (a)).

### 3.2 Constraint for Intrinsic Parameters and Distortion of the Camera Lens

The method can be extended for intrinsic parameters and the distortion of the camera lens. Suppose that the distortion of the camera lens is in proportion to the cubed range from the center of the image. In this case, Eq. (1) is given by

$$u = \alpha_u \frac{X}{Z}\left(1 + k\frac{X^2+Y^2}{Z^2}\right)$$

$$+s\frac{Y}{Z}\left(1 + k\frac{X^2+Y^2}{Z^2}\right) + u_0, \tag{7}$$

$$v = \alpha_v \frac{Y}{Z}\left(1 + k\frac{X^2+Y^2}{Z^2}\right) + v_0, \tag{8}$$

where $k$ is the distortion parameter of the camera lens. With the same procedure given above, we can obtain the linear equation camera parameters and the distortion of the camera lens. Therefore, they can be calculated by obtaining six or more points and applying the simple linear least squares method.

The parameters for the 3D-2D registration can be obtained by the proposed linear equation in a case in which intrinsic parameters or the distortion of the camera lens is either given or not (see Section 2).

## 4 Image Modification for Reducing False Matches

A range intensity and color images differ, and the pair of matched feature points usually contains a number of false matches. Incorrect candidates are extracted from a range intensity image for the following reasons.

- In a range intensity image, a boundary region of a target object has no texture
- The S/N ratio of a range intensity image is low

Therefore, a range intensity image is made to be similar to a color image to reduce false matches before using SIFT.

(a) Color image (b) Background image (c) Range intensity image (d) Combined image

**Fig. 3.** Combined range intensity image with a background image of a color image

### 4.1 Combined Range Intensity Image with a Background Image of a Color Image

When the features of a range intensity image are extracted using SIFT, as a boundary region of a target object has no texture, a number of false matches are extracted. Therefore, it is needed to solve the problem which is a boundary region of a target object has no texture in a range intensity image. For reducing false matches, a range intensity image is combined with the grayscale image which is transformed from a background image of a color image. To combine two images, the 3D information of a projected range intensity image is used. Fig. 3 (a) is an example of a color image. Fig. 3 (b) shows a background image of a color image, Fig. 3 (c) shows a range image, and Fig. 3 (d) illustrates the range intensity image which is combined with a background image.

In addition to matching two images, features that are extracted in a background image are eliminated.

### 4.2 Correction of Range Intensity Image

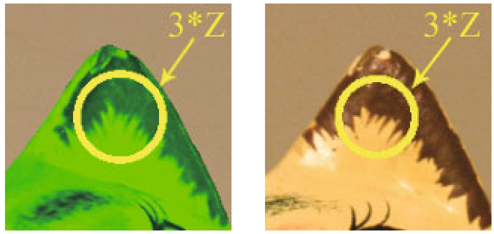
A range intensity image is affected by the following factors.

– Distance to each measured point
– Normal vector at the measured point
– Sensor-specific characteristics

By the factors above, to match a range intensity image and a color image, incorrect candidates are extracted from a range intensity image. Therefore, instead of a raw range intensity image, we used a corrected one that was obtained by the method proposed by Shinozaki et al. [13].

## 5 Soft Matching

To remove false matches, if a simple robust estimate, RANSAC [12], is used, correct matches can be removed. Therefore, soft matching is used before using RANSAC. First, a small difference in the scale of the correct matches is used. Then, the appearance of the pair of matched feature points is used. We use Bhattacharyya distance to evaluate the similarity of the appearance.




(a) Range intensity image (b) Color image

**Fig. 4.** Area of calculating the Bhattacharyya distance ($z$ is the scale information from SIFT)

### 5.1 Scale Information from SIFT

First, false matches were removed by using scale information from SIFT. If the pair of matched feature points is correct, the difference in the scale of each feature point will be small. Therefore, false matches are removed by the following threshold processing.

$$\begin{cases} if \quad (|x| < \mu_x - 0.8 * \sigma_x) : correct \\ otherwise \qquad\qquad\quad : false \end{cases} \tag{9}$$

where $x$ is the difference of scale in each feature point, $\mu_x$ is the average of $x$, and $\sigma_x$ is the standard deviation of $x$. The constant 0.8 was determined empirically.

### 5.2 Matching Reliability from the Bhattacharyya Distance

If the pair of matched feature points is correct, the intensity of the area of each feature point will be similar. We use Bhattacharyya distance for calculating the similarity. The Bhattacharyya distance $S$ is given as

$$S = \sum_{u=1}^{m} \sqrt{p_u q_u}, \tag{10}$$

where $p_u$ and $q_u$ are the two normalized color histograms made from hue information, $u$ is the hue number, and $m$ is the number of the elements of the hue. We propose a matching reliability using Bhattacharyya distance of the intensity of area of the pair of matched feature points. The matching reliability is defined as

$$p = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(1-S)^2}{\sigma^2}\right). \tag{11}$$

This equation represents a normal distribution with the average zero and the standard deviation $\sigma$. In this paper, $\sigma$ is the standard deviation of $S$, and the greater the Bhattacharyya distance, the higher the matching reliability.

It is important to determine the area for calculating the Bhattacharyya distance. To determine the area appropriately, scale information from SIFT is used. In the previous chapter, the difference in scale of each feature point is small. The

Bhattacharyya distance is then calculated $(3.0 * z)$ from SIFT keypoints (see Fig. 4 ), where $z$ is the scale from SIFT and $(3.0 * z)$ is the area of description of SIFT.

The method for removing false matches is shown below. First, the Bhattacharyya distance $S$ is calculated by using Eq. (10). Secondly, the matching reliability $p$ of the pair of matched feature points is calculated by using Eq. (11). False matches are then removed by the following threshold processing.

$$\begin{cases} if \quad (p > \mu_p - 1.0 * \sigma_p) : correct \\ otherwise \qquad\qquad\quad : false \end{cases} \tag{12}$$

where $\mu_p$ and $\sigma_p$ are the average of matching reliability $p$ and the standard deviation of matching reliability $p$ respectively. In addition, the constant 1.0 is determined empirically.

Finally, after soft matching is used, RANSAC is applied. In this stage, its threshold is kept high.

## 6 Experimental Results

The effectiveness of the proposed method is illustrated by experiments with real world objects. These experiments evaluate three points: accuracy, processing time and degree of viewpoint change.

### 6.1 Experiment of System and Initial Values

The range images and range intensity images are obtained with the ShapeGrabber range sensor (scan head SG-102 on a PLM300 displacement system) [14].

A Nikon D70 provides the color images in RAW format. As the laser color is red, the R-channel of the color image is used for registration. The size of the color images is $3008 \times 2000$. The initial value of extrinsic parameters is given as follows. $R$ was set to the unit matrix. $t_x, t_y$ were set as the center of the gravity of the target range. $t_z$ was roughly given manually.

The initial value of intrinsic parameters $a_u, a_v, u_0, v_0$ was obtained using the camera calibration method proposed by Zhang [15]. $a_u, a_v$ were set to 8032 and 8013, respectively. $u_0, v_0$ were set to 1648 and 1069, respectively.

Initial values of other parameters were given as follows. Skew $s$ was set to 0. $k$ was set to zero, i.e., without any distortion of the camera lens.

### 6.2 Making a 3D Geometric Model with Range Intensity Images

The object of registration is Fig. 1 (cat). The dimensions of the object are $w59mm \times h112mm \times d32mm$. Fig. 5 (a) illustrates a 3D geometric model that is integrated from a number of range images and corrected range intensity images with PolyWorks [16]. The number of images is 15, and that of points is 183016 in the 3D geometric model.

(a) cat (b) shiny crane

**Fig. 5.** Geometric model with intensity information

### 6.3 Registration Result

Fig. 6 (a), (b) illustrates the result of the registration. The bright (green) images are the range intensity image, and the dark (red) images are the color image. The result of camera parameters and the distortion of the camera lens are shown below.

$$\mathrm{R} = \begin{bmatrix} 0.883 & 0.037 & -0.468 \\ -0.027 & 0.999 & 0.033 \\ 0.469 & -0.018 & 0.883 \end{bmatrix}, \mathbf{t} = \begin{bmatrix} 151.5 \\ -19.6 \\ -27.3 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_u \ \alpha_v \ s \end{bmatrix} = \begin{bmatrix} 6091.7 \ 6097.6 \ -12.8 \end{bmatrix}$$

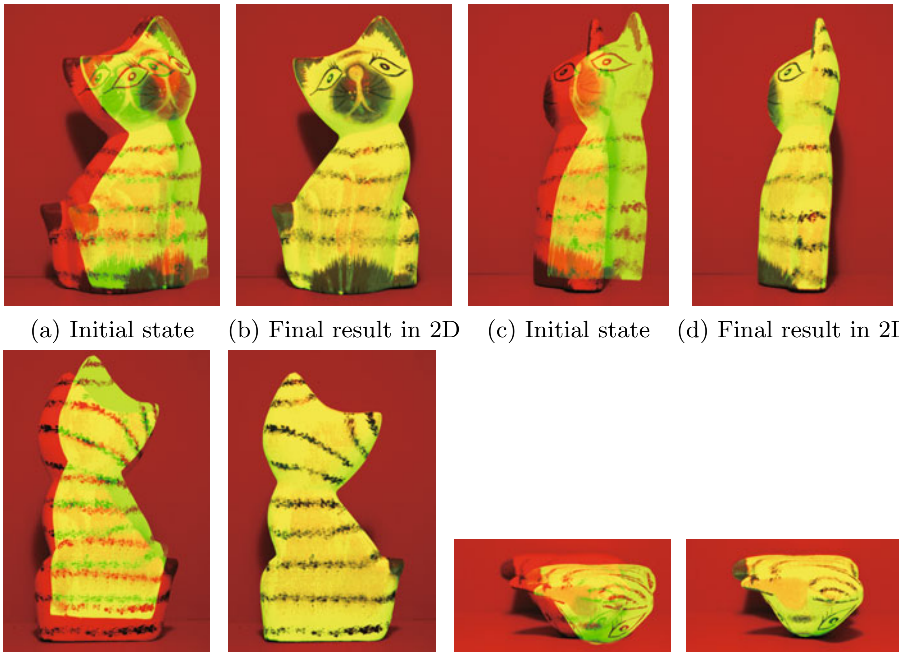
$$\begin{bmatrix} u_0 \ v_0 \end{bmatrix} = \begin{bmatrix} 1362.9 \ 976.7 \end{bmatrix}, k = 0.1892$$

The processing time was about 30.4 s when we used a PC (CPU: Core i7 2.93GHz) with GPU (GeForce GTX260). The correlation is used to evaluate the convergence of the registration.

If the correlation is decreased after several iterations, the stage is updated. In the next stage, the parameters are used, obtaining the highest correlation in the previous stage. The number of iterations is 17 in three stages and the correlation coefficient is 0.8766. Then, Fig. 7 illustrates the result of registration for a variety of viewpoint changes.

This result shows that precise registration can be obtained even when the initial parameters are far from the final estimated values. In addition, by obtaining the extrinsic and intrinsic parameters and distortion of the camera lens, precise registration can be obtained. Fig. 6 (c), (d), (e), (f), (g), (h) illustrate the result of the registration of other angles in the same object (Fig. 6 (a), (b)). The correlation coefficients are $0.8286, 0.8467$, and $0.5467$, respectively. These results show that precise registration can be obtained when a color image of varied angles is used. Fig. 6 (g), (h) show that our method can be applied when the object contains few features.

Fig. 8 (a), (b) show another model (shiny crane). The dimensions of the object are $w44mm \times h175mm \times d64mm$. Fig. 5 (b) shows the 3D geometric model. The number of range and corrected range intensity images is 87 and that of points is 409057 in 3D geometric model. Fig. 8 (c), (d) illustrate the result of



(a) Initial state (b) Final result in 2D (c) Initial state (d) Final result in 2D

(e) Initial state (f) Final result in 2D (g) Initial state (h) Final result in 2D

**Fig. 6.** Registration result: cat

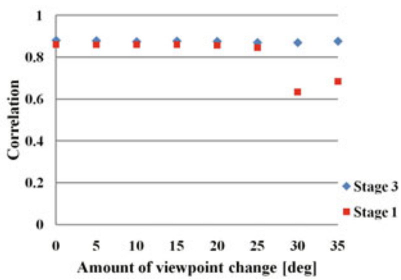



**Fig. 7.** Effect of viewpoint change

the registration. The result of the camera parameters and the distortion of the camera lens are shown below.

$$\mathrm{R} = \begin{bmatrix} 0.990 & 0.378 & -0.137 \\ -0.025 & 0.995 & 0.093 \\ 0.140 & -0.088 & 0.986 \end{bmatrix}, \mathbf{t} = \begin{bmatrix} -31.8 \\ 13.0 \\ -409.9 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_u \; \alpha_v \; s \end{bmatrix} = \begin{bmatrix} 7226.4 \; 7185.8 \; -47.0 \end{bmatrix}$$

$$\begin{bmatrix} u_0 \; v_0 \end{bmatrix} = \begin{bmatrix} 1511.3 \; 1423.2 \end{bmatrix}, k = 0.000$$

(a) Range intensity image (b) Color image (c) Initial state (d) Final result in 2D

**Fig. 8.** Range intensity image and color image and registration result: shiny crane

(a) cat (b) shiny crane

**Fig. 9.** Constructed 3D models with color information

The processing time was about 24.2 s. The number of iterations is 11 in three stages and the correlation is 0.7343.

Fig. 9 (a) illustrates the final result of texture mapping by using five color images. The images were taken from front (Fig. 6 (a)), left (Fig. 6 (c)), back (Fig. 6 (e)), top (Fig. 6 (g)) and left. Fig. 9 (b) illustrates another result of texture mapping with five images. Fig. 8 (b) is the one of the five.

## 7 Conclusions

We have proposed a method for 3D-2D registration using SIFT and the range intensity image. In our approach, the extrinsic and intrinsic parameters and distortion of the camera lens can be obtained simultaneously. Then, we have proposed a method to reduce false matches and soft matching. In soft matching, false matches are weakly removed while correct ones are kept. Before using SIFT, to reduce false matches, a range intensity image is combined with a background image of a color image and corrected. Moreover, we have proposed the soft matching. We achieve precise automatic registration. In future work, a more detailed quantitative evaluation of the method will be performed.

## References


1. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital Michelangelo project:3D scanning of large statues. In: SIGGRAPH 2000, pp. 131–144 (2000)
2. Iwakiri, Y., Kaneko, T.: PC-based realtime texture painting on real world objects. In: Proc. Eurographics 2001, vol. 20, pp. 105–113 (2001)
3. Lensch, H.P.A., Heidrich, W., Seidel, H.P.: Automated texture registration and stitching for real world models. In: Proc. Pacific Graphics 2000, pp. 317–326 (2000)
4. Lavallee, S., Szeliski, R.: Recovering the position and orientation of free -form objects from image contours using 3D distance maps. IEEE Trans. Pattern Anal. Mach. Intell. 17(4), 378–390 (1995)
5. Neugebauer, P.J., Klein, K.: Texturing 3D models of real world objects from multiple unregistered photographic views. In: Proc. Eurographics 1999, pp. 245–256 (1999)
6. Boughorbel, F., Page, D., Dumont, C., Abidi, M.A.: Registration and integration of multi-sensor data for photo-realistic scene reconstruction. In: Proc. Applied Imagery Pattern Recognition, pp. 74–84 (1999)
7. Umeda, K., Godin, G., Rioux, M.: Registration of range and color images using gradient constrains and range intensity images. In: Proc. of 17th Int. Conf. on Pattern Recognition, vol. 3, pp. 12–15 (2004)
8. Kurazume, R., Nishino, K., Zhang, Z., Ikeuchi, K.: Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute. In: Proc. Fifth ACCV, pp. 99–106 (2002)
9. Elstrom, M.D., Smith, P.W.: Stereo-based registration of multi-sensor imagery for enhanced visualization of remote environments. In: Proc. of the 1999 Int. Conf. on Robotics Automation, pp. 1948–1953 (1999)
10. Boehm, J., Becker, S.: Automatic Marker-Free Registration of Terrestrial Laser Scans using Reflectance Features. In: 8th Conf. on Optical 3D Measurement Techniques (2007)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60(2), 91–110 (2004)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 16(24), 381–395 (1981)
13. Shinozaki, M., Kusanagi, M., Umeda, K., Godin, G., Rioux, M.: Correction of color information of a 3D model using a range intensity image. Comput. Vis. and Image Understanding 113(11), 1170–1179 (2009)
14. ShapeGrabber, `http://www.shapegrabber.com`
15. Zhang, Z.: A flexible new techniques for camera calibration. IEEE Trans. Pattern Anal. Martch. Intell. 22(11), 1330–1334 (2000)
16. PolyWorks, `http://www.innovmetric.com`