Three-dimensional shape model reconstruction from multiple-view range images and color images

Yoshida, Kazushi, Saito, Hideo

Kazushi Yoshida, Hideo Saito, "Three-dimensional shape model reconstruction from multiple-view range images and color images," Proc. SPIE 4572, Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision, (5 October 2001); doi: 10.1117/12.444175

Event: Intelligent Systems and Advanced Manufacturing, 2001, Boston, MA, United States

Three-dimensional shape model reconstruction from multiple view range images and color images

Kazushi Yoshida and Hideo Saito†

Department of Information and Computer Science of Keio university
3-14-1 Hiyoshi Kohoku-ku, Yokohama, 223-8522, Japan

ABSTRACT

Recently, there is an increasing interest in capturing 3D models of real objects. The range scanner can acquire high quality shape data of the object, but the texture image for surface rendering obtained by the scanner is not generally high resolution and high quality. High-resolution color images at different position are generally taken in addition to the range data so that more realistic images can be rendered from the captured 3D model using such high quality textures. We propose the method of modeling high quality 3D model in shape and appearance by aligning multiple view range images obtained by a range scanner and multiple view color images taken by a digital camera around the object. Color images used as textures are calibrated by Tsai's method, in which lens distortion is also calibrated. On the other hand, a surface model of the object is created by registering and integrating range data sets taken from multiple directions. For registration, we use color ICP (Iterative Closest Point) algorithm that aligns two surfaces using color images and surface shape of the object. For integration, we build voxel model from range images and then detect polygons by Marching Cubes algorithm. We apply textures of high-resolution images to the surfaces by blending, and finally reconstruct realistic 3D model concerned with shape and appearance.

Keywords: 3D reconstruction, multiple view range images, ICP algorithm, texture mapping

1. INTRODUCTION

There is an increasing interest in modeling real objects in the areas of business, education, and entertainment. Capturing 3D models is important especially in interactive applications, 3D games, movies, and preservation of historical architectures or works of art. To create 3D model, we generally need shape and texture information of the object. When we input such information into computers manually, it is troublesome and wastes much time. If we measure the shape information of the object and create a model using the information, we can decrease labor and create more realistic 3D model. The method of shape measuring is most important subject in the recent computer vision field. On the other hand, we can get shape and texture information using some kinds of range scanners. However a texture image obtained by range scanner doesn't usually have enough resolution to create realistic rendered images by the range data. In the most cases, we should take high resolution images of the object using a high-resolution digital camera, which is used as high quality textures in place of the low-resolution textures provided by range scanner. However, it is difficult to accurately set positions of the devices (range scanner and camera). Therefore, it is important to register the high quality texture images taken by the camera with range data taken by the range scanner. Such accurate registration is significant for creating more realistic 3D model.

In this paper, we present a method of modeling high quality 3D model in shape and appearance (color, texture) by aligning and merging multiple view range images obtained by range scanner and high resolution color images obtained around the object. To create real models, a single view data cannot completely convey the shape of the scene, thus it is necessary to merge multiple view shape information. Like many multi-view merging algorithm, our approach has two components. Overview is shown in Fig.1. First step is registration stage that aligns multiple view range images and high-resolution color images. To align range data sets, we use the modified iterative closest point (ICP) algorithm[1 2 3] that aligns two surfaces using color and shape information. To align range data and high-resolution color images, we use Tsai's calibration method[5 6] that is generally used for calibrating camera with lens

E-mail address
Kazushi Yoshida: kazushi@ozawa.ics.keio.ac.jp
Hideo Saito: saito@ozawa.ics.keio.ac.jp
†Telecommunications Advancement Organization of Japan (TAO)

distortion. After registration, there is an integration stage that merges multiple view data sets into single seamless model using position and surface information of range data sets and then mapping textures of high-resolution images onto the model. There are many researches about integration of range data sets[7 8].[9] We merge range data sets inside voxel grid space, and then detect meshes by Marching Cubes Algorithm.[10] Applying textures of high-resolution color images by blending to the surface model, we create the final realistic model which is high quality in shape and appearance.

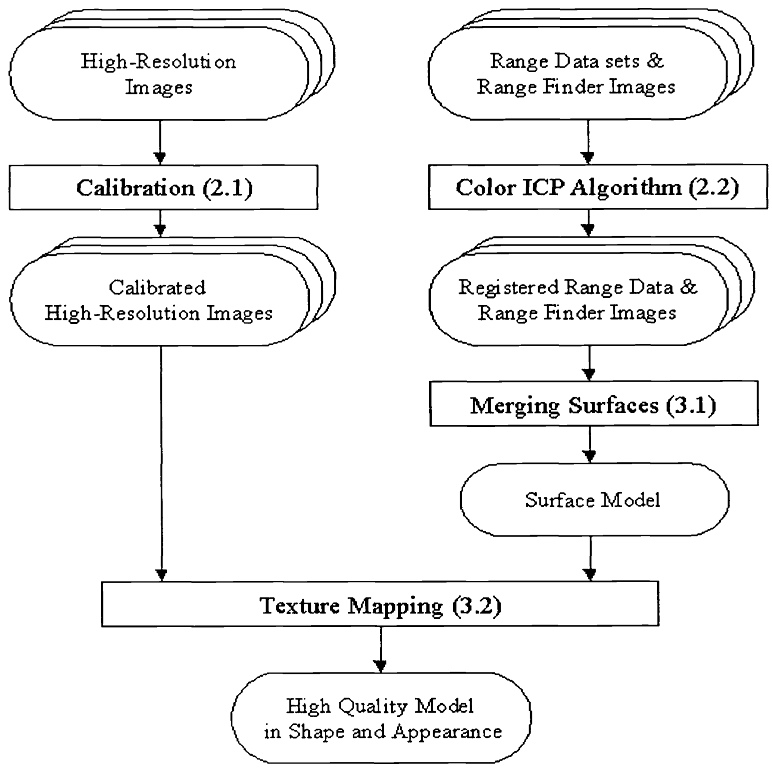


Figure 1. Overview of proposed method.

2. REGISTRATION

In the registration stage, two kinds of registration are performed. First is registration between high-resolution color images taken by a digital camera and range data sets taken by a range scanner. Second is registration of range data sets of the object taken from different angles. In our method, we assume that there is no information about the position of the camera and the range scanner, so that just capturing the object can be sufficient for model generation. Therefore, the processes of registration are very important because inaccuracy of alignment results in the error appearance of final model created by merging all of the data sets (range data and high resolution color images). We assume to use a range scanner that provides not only range data of the object but also intensity images of the object, which is not sufficient resolution for realistic rendering. In this article, we express the intensity image taken by range scanner as "range scanner images", while we call high-resolution images taken by high-resolution digital camera as "high-resolution images".

2.1. Calibration of High Resolution Images

In our method we have no restriction about the position of acquiring all of the data sets, there is no information about camera positions of high-resolution images related to range data or world coordinate system. To project the textures of these high-resolution images onto the object surface model, it is necessary to get the geometrical information of images related to data sets of world coordinate system. In our method, we calibrate high-resolution images by Tsai's calibration method. Tsai's camera model is pinhole camera model and considers parameters inside camera (focal length f, radial lens distortion *kappa*, scale factor S_x, pixel size (d_{px},d_{py}), lens center (C_x,C_y), and extrinsic camera parameters (rotation and translation parameters about x, y, z axis (R_x, R_y, R_z, T_x, T_y, T_z)). These parameters represent the relation of world coordinate and image coordinate of high-resolution color images. If we can avoid the radial lens distortion, we show the projection matrix from world coordinate into image plane as expressed in the next equation (1).

$$\begin{bmatrix} H \cdot x_i \\ H \cdot y_i \\ H \end{bmatrix} = \begin{bmatrix} f \cdot S_x/d_{px} & 0 & C_x \\ 0 & f/d_{py} & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

Where (X_w, Y_w, Z_w) shows world coordinate point and (x_i, y_i) is image plane coordinate, R_{11} to R_{34} are determined by R_x, R_y, R_z. This matrix doesn't include radial lens distortion parameter *kappa*. To reflect the lens distortion parameter, we should take next relations (2)(3) into account.

$$x_u = x_d \times (1 + kappa \times ({x_d}^2 + {y_d}^2)) \tag{2}$$

$$y_u = y_d \times (1 + kappa \times ({x_d}^2 + {y_d}^2)) \tag{3}$$

(x_d, y_d) is the image plane coordinate concerning lens distortion and (x_u, y_u) is the distortion free coordinate. By considering these relationships, we can get more accurate geometrical information of high-resolution color images. In our method we give matching points of range scanner images and high-resolution images more than 7 points by hand to get 3D-2D information. We calculate the most suitable camera parameters using this correspondence map information.

2.2. Registration of Range Images

To construct 3D model from multiple range data sets taken from different view directions, it is necessary to align all data sets in a common world coordinate system. In this method, we have no restriction about the position of acquiring range data sets, therefore we give some correspondence points to determine rough initial positions of range data sets. Then we determine rigid transformation of them by Color ICP algorithm.[4] ICP algorithm is popular method for registration. Color ICP algorithm is the extended method of ICP to improve the accuracy of registration. We apply Color ICP because data points obtained by range scanner also have color information as range scanner images.

2.2.1. Color ICP algorithm

After computing coarse position of each range data, the data sets are exactly aligned by Color ICP algorithm. We remark only one data set and adjust coordinate transformation. Iterating this process for all data sets, then the whole registration is finally achieved.

Color ICP method is as follows. Among all points P_i of remarked range data, the closest point C_i is found. All the pair points (P_i, C_i) are checked, and then the distance d_i of the pair points is calculated and summed up to D_i. Transformation parameters that provide minimum D_i are found by an optimization algorithm. The transformation parameters are 6 parameters about rotation and translation along Euclid axes (X, Y, Z).

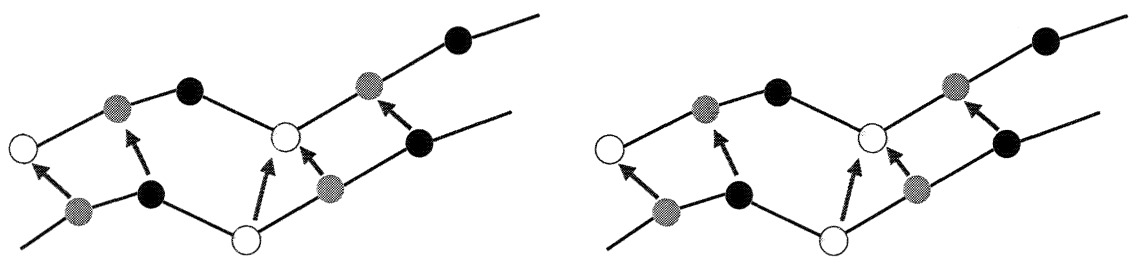
If optimal transformation is determined, then the closest point is sought again, and the seeking is repeated until the distance D_i between closest points falls below a threshold. Then next range data is processed in the same way. After all the range data sets were regarded, the area where closest point was sought for is narrowed and retry about all range data again, and the seeking is repeated until the convergence criterion is met.

2.2.2. Judging A Closest Point

Since depth and color intensity are obtained by range scanner, more precise correspondence points can be detected by considering not only the position in three dimension but also the color value in RGB. In our method, Color ICP algorithm decides closest points considering 6 parameters of the position and the color. The distance of two points p_1, p_2 in 6 dimension is defined as follows.

$$d(p_1, p_2) = (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \\ + \alpha_1 (R_1 - R_2)^2 + \alpha_2 (G_1 - G_2)^2 + \alpha_3 (B_1 - B_2)^2 \quad (4)$$

where (X_n, Y_n, Z_n) is 3D coordinate and (R_n, G_n, B_n) is the color of the point, and α_n shows weight coefficient. If no color information was used in registration, closest point is simply determined by Euclidean distance between the points. In this case, if initial position is located as shown in Fig.2(a), the transformation converges to wrong location. On the other hand, more accurate registration can be obtained using color information as shown in Fig.2(b).



(a) Normal ICP. (b) Color ICP.

Figure 2. An advantage of Color ICP.

3. INTEGRATION

After all data sets were aligned, the next stage is integration stage, in which all aligned data sets were merged into a single seamless model. In this stage, a single surface model is first built using surface information of all range data sets. Next all aligned high-resolution images are applied to the model, where the overlapping textures are blended, and finally high quality model in shape and appearance is completed.

3.1. Merging Surfaces

Many approaches are researched about integration algorithm. Some of the algorithm construct 3D implicit function describing the surface using voxel data structure and then it is converted to a surface model using the Marching Cubes algorithm.[10] The important thing in these methods is how the implicit surface is constructed and the volumetric data is organized. We use the Consensus Surface algorithm[9] that is very effective when handling noisy data. We organize the implicit function using this method and then detect meshes and convert into surface model using Marching Cubes algorithm.

3.1.1. Consensus Surface Method

Overview of Consensus Surface method is as follows. Detail is well acquainted in reference. Firstly we define voxel grid region which includes all range data sets. Then signed distance is calculated for all voxels. The procedures are shown as next.

Algorithm $ConsensusSurfaceMethod$
Input: voxel point $\mathbf{x}$
Input: triangle set $\mathbf{T}=\bigcup_i T_i$ (i : range data number)
Output: the signed distance d
1. $\langle p, n \rangle \leftarrow ClosestConsensusSurface(\mathbf{x}, T_i)$
2. $d \leftarrow ||x - p||$
3. if $(n \cdot (x - p) > 0)$
4. then $d \leftarrow -d$
5. return d

$ClosestConsensusSurface(\mathbf{x}, T)$ computes the location of the closest Consensus surface point p and it's normal n from remarked voxel coordinate. Consensus surface is virtual surface determined by weight average of coordinates and normals of range data sets located near the voxel and high possibility that real surface exists. Absolute value of returned d is the distance to nearest consensus surface and the sign determined by dot product with the normal of nearest consensus surface means whether voxel is likely to be included inside the object or not.

Computing in this manner, the voxel value near which real surface probably exists is theoretically zero and higher value heading for inside the object and lower value for outside it adversely. The important thing for making implicit function is the location where voxel's sign changes and magnitude of these voxels. The implicit function composed in this way is converted to a surface model of triangle meshes by Marching Cubes algorithm. Generally the time cost of merging range data is very long, so some refinements are expected. Our method previously computes the projection matrix of each range image for it's image plane and project remarking voxel onto image planes of all range images so as to figure out the area of seeking closest point to reduce computing time.

3.1.2. Marching Cubes Algorithm

In this algorithm, triangle meshes are detected from volumetric data. Outline is as follows. We handle at one time cubic area composed of 8 voxels that are neighboring each other. In our 3D implicit function representation, the value of voxel is plus where inside the surface and minus where outside the surface. Surfaces determined by all range data sets will exist boundary of the neighboring voxels where the sign of voxels changes. So triangle meshes are formed just 15 patterns related to the sign of the 8 voxels. Iterating this process for all areas of the voxel grid can convert 3D implicit representation into single surface model. The vertices of a triangle mesh are decided in consideration of the weight of the neighbor voxels' values. The smooth surface is formed when the weight values of voxels are well computed.

3.2. Texture Mapping

To apply textures of calibrated high-resolution images onto surface mesh model, visibility of each mesh from viewpoint of camera must be determined. Our method determines the visibilities using Z-buffer. Some triangle meshes have number of candidate high-resolution images to be rendered, because multiple cameras can view them. In this case, we can adopt only one texture to a mesh considering the area of triangle in texture image, or we can also blend some texture images. The former method results in brilliant texture because original high-resolution is retained, but discontinuity appears if next texture is applied from other camera image. The latter method results in uniform model with smooth appearance in texture boundaries but less clear texture. Our method is the midpoint of them. If mesh is caught by some texture images, we select the best and second texture considering the area of triangle in texture image, then apply by blending. If there is a mesh which no texture image captures, the texture of neighboring meshes are applied to it.

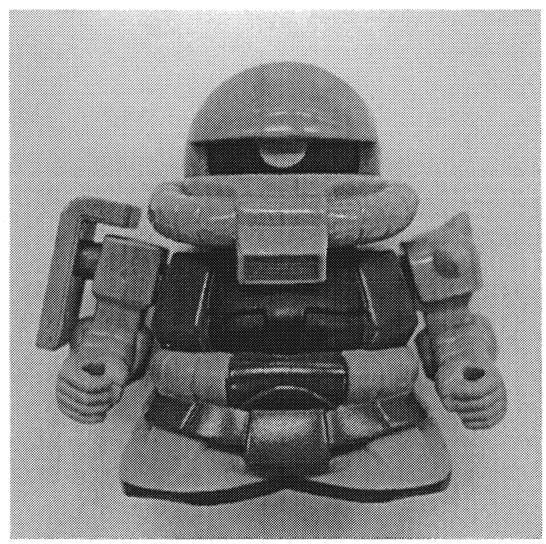
High-resolution images are well calibrated because of manual input of corresponding points. However, there often arises the error between range data sets and merged surface model. This error causes the problem that background area of high resolution images are applied to the surface. Our method avoids the problem separating object and background area loosely in advance by projecting all range data sets onto the image. If a mesh is projected into boundary area of them, the mesh is judged invisible from the viewpoint of the image. This avoids mapping background onto surface.

4. EXPERIMENTAL RESULTS

Here we present some experimental results of our implementation of the 3D modeling algorithm described in this paper. The range scanner we use in this experiment can measure 400 $\times$ 400 size range images with each pixel containing a 3D coordinate and texture images as same size. The high-resolution images were 1600 $\times$ 1200 size and captured by a normal digital camera. For our experiments, we selected two objects to model using our algorithm; a cellular phone and a toy plastic-made robot. Either of both is reconstructed by 8 range images and 8 high-resolution images. The number of the voxel grid used to merge range images was 128 $\times$ 128 $\times$ 128. Results of registration of range data sets, texture mapping of range scanner images, and texture mapping of high resolution images in our method is shown in Fig. 4 $\sim$ Fig. 9.

Fig.6, 9 shows that each model applied with high-resolution images as textures is superior in quality than applied with range scanner images in Fig.5, 8. Our method achieved good results that high-quality model about both shape and appearance was created. However, some regions where textures were out of alignment and shapes of the models lacked accuracy found in places. These areas were generated by errors of calibration of high-resolution images and registration of range images. Reducing these areas contributes to more realistic models. We must tackle with these error problems in future.

(a) A cellular phone.



(b) A toy robot.

Figure 3. Physical objects used in experiments.

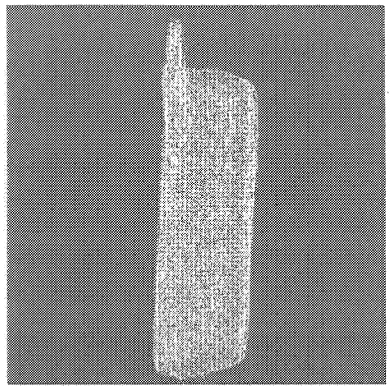
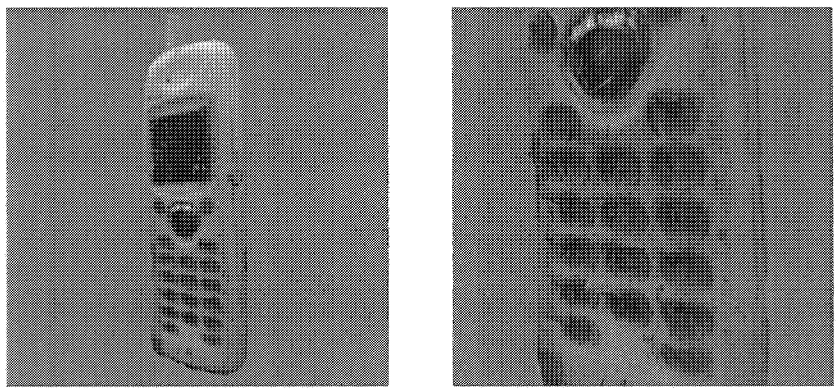
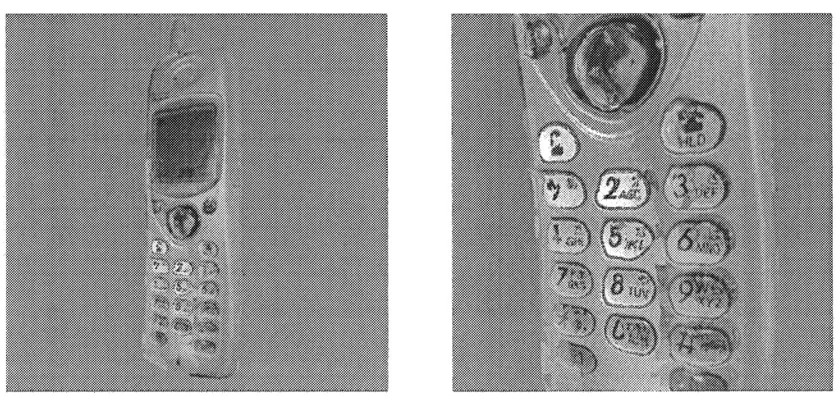


Figure 4. Registration of range images (cellular phone).



(a) Virtual view1. (b) Virtual view2 (close up).

Figure 5. Texture mapping of range scanner images (cellular phone).



(a) Virtual view1. (b) Virtual view2 (close up).

Figure 6. Texture mapping of high-resolution images (cellular phone).

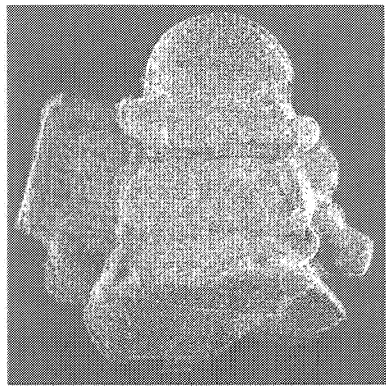
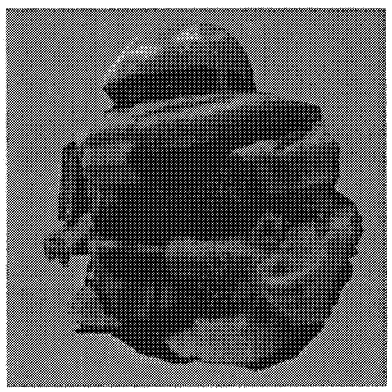


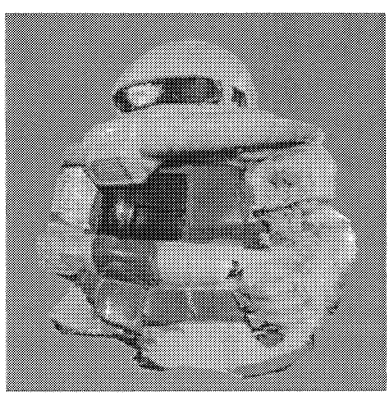
Figure 7. Registration of range images (toy robot).



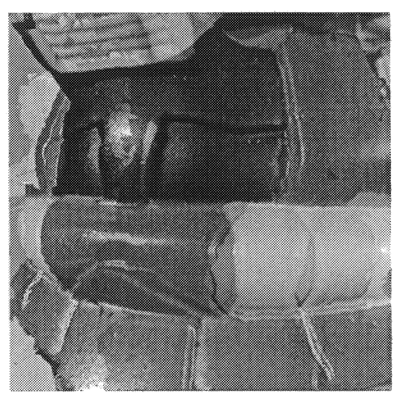
(a) Virtual view1.

(b) Virtual view2(close up).

Figure 8. Texture mapping of range scanner images (toy robot).



(a) Virtual view1.



(b) Virtual view2(close up).

Figure 9. Texture mapping of high-resolution images (toy robot).

5. CONCLUSION AND FUTURE WORKS

We propose the method of modeling real object using multiple view range images and high-resolution images. To create high quality model in shape and appearance, we register range images obtained by range scanner and create seamless surface model, then apply textures of high resolution color images captures at unrelated positions.

We can create realistic model in appearance using high-resolution textures obtained independently with range scanner, as we said in experimental results, accuracy of registration of range images can still be improved. So we should take ourselves to tackle with this registration problem. Reducing registration errors will improve the quality of the created model. Besides, now calibration of high resolution images are done by inputting correspondence points manually. We will also develop automatic correspondence points detection for labor saving in future.

REFERENCES

1. P. Besl, N. McKay, "A method of registration of 3-D shapes," *IEEE Trans. PAMI*, vol. 12 (2), pp. 239-256 (1992).
2. Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp.145-155, (1992)
3. Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Computer Vision*, vol. 13 (2), pp. 119-152, (1994)
4. Andrew Edie Johnson, Sing Bing Kang, "Registration and integration of textured 3D data," *Image and Vision Computing*, vol. 17, pp. 135-147, (1999)
5. Roger Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, FL, pp. 364-374, (1986)
6. Roger Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, August pp. 323-344, (1987)
7. Y. Chen and G. Medioni, "Surface description of complex objects from range images," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 153-158, (1994)
8. B. Curless and M. Levoy, "A volumetric method for building complex models from range images," *Proc. SIGGRAPH*, pp. 303-312, (1996)
9. M. D. Wheeler, Y. Sato, K. Ikeuchi, "Consensus surfaces for modeling 3D objects from multiple range images," *Proc. Image Understanding Workshop*, pp. 911-920, (1997)
10. W. E. Lorensen, H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *Computer Graphics*, vol. 21, No. 4, pp. 163-169, July (1987)