Object Shape Reconstruction and Pose Estimation by a Camera Mounted on a Mobile Robot

Kimitoshi Yamazaki, Masahiro Tomono, Takashi Tsubouchi and Shin’ichi Yuta
Intelligent Robot Laboratry, Institute of Engineering Mechanics and Systems University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan
Email: {yamazaki, tomono, tsubo, yuta}@roboken.esys.tsukuba.ac.jp

***Abstract*— This paper describes how to acquire 3-D shape of an object and to estimate robot pose by observing an unknown object. A single camera is placed on the robot, images of the target object are captured in streams by the camera on a mobile robot moving around the object. Computer vision techniques are utilized under the condition that shape, size and accurate position information of the target object are not given in advance, but the robot finds the target object and reconstructs the 3-D object shape by itself. In addition, it is necessary that the robot grasps relative pose between itself and the target object. Experimental results show the effectiveness of our method in the point of robustness and accuracy.**

I. Introduction

This paper describes how to acquire 3-D shape of an object and to estimate robot pose by observing an unknown object through captured images. The images are obtained by a camera on a mobile robot moving around the object.

Manipulation of objects in realistic world such as object picking up by a mobile manipulator provides our motives to study entitled issues. Conventionally, object manipulation by a manipulator with an end effector has been performed on a fixed platform, where the number of reachable objects to be manipulated is limited. However, once the manipulator gets mobility in realistic environment, the number of reachable objects may extremely increase. It becomes difficult to hold the shape information of all of the objects in the environment. Necessity of shape acquisition ability through vision arises from such a situation. Artificial marks on the objects will be helpful to reduce the computational complexity for feature point identification[9]. A part of shape knowledge could be due to identify the object especially in Bin-Picking problems[6].

We prefer to take such an approach that shape, size and accurate position information of the target object is not given in advance, but the robot finds the target object and reconstructs the 3-D object shape by itself. In addition, it is necessary that the robot grasps relative pose between itself and the target object. In these reason, we develop a framework for acquisition of the object shape and the robot pose fast and accurately.

This paper is organized as follows: In the next Section, related works are described. Section III presents a scheme of our research and problem definition. The outline of acquiring 3-D object shape and robot pose are described in Section IV, and improvement of our proposed method for real robot moving in Section V. Section VI provides experimental results to show effectiveness of our method. Section VII concludes this paper.

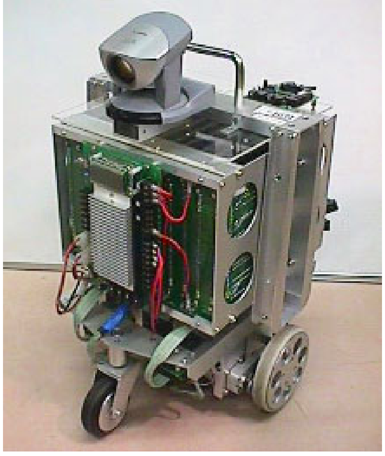


Fig. 1. Wheel Based Mobile Robot

II. Related Works

In several issues, an object in real environment are manipulated by an autonomous mobile robot.

Miura et al.[4] realized that a mobile manipulator as a service robot grasps a PET bottle and brings it to a person. In this research, object models based on image data are given in advance, and the robot recognizes the objects through the models. Meanwhile, there are several research issues for object handling. For example, autonomous mobile robot navigation which includes Door Opening behavior[5], or object picking and carrying the specific object[2] were realized. These are a sort of research where the robot performs planning and manipulation based on the knowledge about its target object, whose models are given in advance.

On the other hand, we focus on acquiring the object model autonomously in this paper, however any object models will not be provided in advance. Our approach is to use several images which are captured around the object. After the object models recognition is realized, we will proceed to object handling by a mobile manipulator.

A large number of issues to reconstruct 3-D shape of the object from several images are contributed in the field of Computer Vision. Especially, SFM (Structure From Motion) approach are well-known way to acquire camera motion and 3-D object shape simultaneously under the condition that no or almost no information of camera motion. Factorization[1] and related methods depending on Epipoler Geometry[10] are useful under the condition that correspondence of feature points are known among the images.

Our proposed method in this paper is closely related to [9], because camera pose and object shape are also estimated from image streams in our framework. In this paper, we improved the method [9] in robustness and accuracy. The mobile robot acquires 3-D models of the object and the positions where the robot passes based on our improved method while it moving.

III. Scheme and Problem Definition

In the framework of this paper, a wheeled mobile robot which is equipped with a single pan-tilt type camera (Fig.1) is used. Shape of object is reconstructed in 3-D and simultaneously, the robot(camera) pose is estimated based on image streams which are captured from several viewpoints. In the rest of this paper, we name 'camera pose' as "robot position and camera pose".

A. Target Object Condition

The target object which is observed by the robot is selected among several objects in the environment while the robot is moving. We challenge ourselves to more difficult problem setting than conventional framework.

1) No artificial marks is placed on the object.
2) Not only planer surface object but curved surface object is allowed.
3) Texture on the surface of the object is essential.

The third assumption arises from the fact that texture is necessary to make correspondence of the same feature point on the object in a series of images. We employ KLT-Trakcer[3] to extract and track feature points from images.

We assume that the object is placed on the floor because the camera on the robot (Fig.1) matched the object with tilted angle. However this assumption will not be essential.

B. Provided Conditions for Object Shape Reconstruction

We have provided the conditions for the mobile robot to reconstruct object shape by moving around the object as follows:

- The robot itself select watching object autonomously, while the robot moves in the environment. It is assumed that there are several objects not only the target object because the robot moving in the realistic world.
- Immediate and sequential shape reconstruction is necessary for adaptive motion change for the robot.
- It is only the knowledge given in advance that the target object is placed on the floor. No knowledge about size nor shape of the object is given.

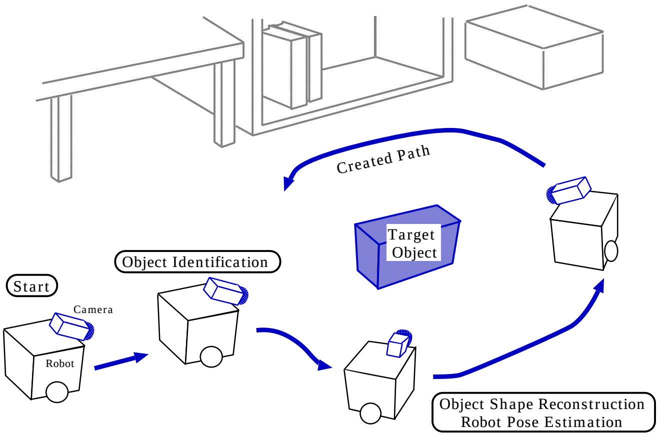


Fig. 2. Movement Flow

C. The Mobile Robot

Odometry (a method of self position estimation by integral of wheel rotation using encoders) and pan-tilt angle data from the camera are the only way to directory obtain the camera pose. In this paper, we call this estimation way 'motion model'. However, the motion model is not accurate enough because of several factors, such as body slip, camera angle error and missed synchronization between the measurement of the motion model and the captured image.

As our purpose is 3-D shape acquisition of the small object, camera pose from the motion model does not have enough accuracy for the proposed purpose. In this reason, our approach to acquire camera pose is divided into two patterns from the condition of captured image streams. If the robot motion is simple, camera pose is estimated by compensation for motion model because it is assumed that motion model is near to real. If the robot motion is complexd, camera pose is estimated by image streams only. In our approach, searching the target object from real environment is performed while the robot moves simple trajectory. Observing the target object intensively and acquiring details of the object is performed while the robot moves around the object with complexd and control of the camera angle.

IV. Simultaneous Camera Pose Estimation and Object Shape Reconstruction

Motion of the robot to achieve camera pose estimation and object shape reconstruction simultaneously is outlined by 3 steps as follows (Fig.2):

1) **motion** : Go straight on the floor.
   mission : Select a target object among several objects existing on the floor.
   （*Target Object Detection*）
2) **motion** : Approach the target object.
   mission : Creating initial camera pose and object 3-D shape from image streams which are captured while the robot approaches the target object.

（*Generation of Initial Camera Pose and Object Shape*）

3) **motion** : Move around the target object.
mission : Estimating camera pose and object shape from each image which is captured while the robot moves around the target object. Camera pose and object shape is obtained simultaneously.
（*Simultaneously Estimation of Camera Pose and Object Shape*）

Fast and accurate enough algorithm is provided in this paper by the authors, where techniques in computer vision such as Motion Stereo, Factoriztion method and Nonlinear Minimization are fully and effectively utilized.

A. Target Object Detection

This step is aimed at selecting a particular object and getting accurate relative position between the camera and the object. Motion stereo and Nonlinear Minimization is utilized in this step.

The robot moves straight while searching the object. This straight motion provides relatively reliable camera pose which is estimated from the motion model of the robot and it is enough to apply Motion Stereo to a pair of images captured at different viewpoints. However, the camera pose from the motion model includes considerable error to obtain accurate pose and 3-D position of the feature points of the object. We apply Nonlinear Minimization to correction of the camera pose and the 3-D feature point positions of the object which are obtained by Motion Stereo. Finally, clustering operation is applied to the obtained feature points and a unique target object is selected.

Motion Stereo is a method which calculates 3-D position of feature points by means of minimization of the linear function as follows:

$$C = \|\mathbf{X} - s_1\tilde{\mathbf{m}}_1\|^2 + \|\mathbf{X} - s_2\mathbf{R}\tilde{\mathbf{m}}_2 + \mathbf{T}\|^2 \ , \qquad (1)$$

where $\tilde{\mathbf{m}}_1$ and $\tilde{\mathbf{m}}_2$ are extended vectors which are the coordinates of an correspondent feature point in image 1 and 2 respectively. $\mathbf{X} = (X, Y, Z)$ is the 3-D position of feature points. $\mathbf{R}$ is relative camera poses and $\mathbf{T}$ is relative position between the two images.

If relative pose between the camera and the robot body is known, $\mathbf{R}$ and $\mathbf{T}$ is obtained from the motion model of the robot. This conventional Motion Stereo based on linear function minimizaition is fast in processing time. However, obtained $\mathbf{X}$ suffer from errors of $\mathbf{R}$ and $\mathbf{T}$ if they are considerable.

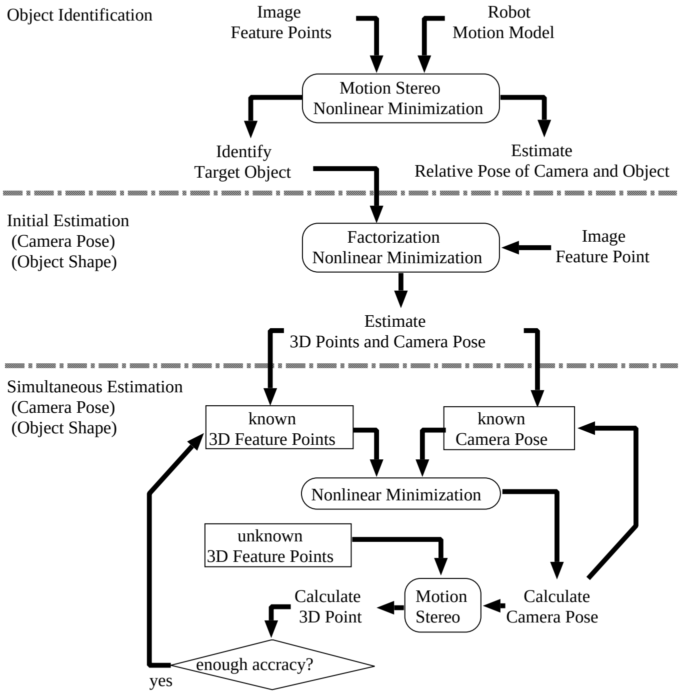


Fig. 3. Algorithm

To cope with the error of camera pose from motion model, Nonlinear Minimization is applied. 3-D position of feature points from Motion Stereo and camera poses from motion model are utilized as initial parameter of this process. Equation (2) is a nonlinear evaluation function. This represents squares of difference between image feature points from KLT-Tracker and reprojection of the obtained 3-D feature points to present image plane. Optimization is realized to find the camera pose and 3-D position of feature points that minimizes Eq.(2):

$$\begin{aligned} C = & \sum_{j=0}^{2}\sum_{i=0}^{P}\left(\frac{\mathbf{r}_{xj}^T\mathbf{m}_{ij}}{\mathbf{r}_{zj}^T\mathbf{m}_{ij}} - \frac{X_i + t_{xj}}{Z_i + t_{zj}}\right)^2 \\ & + \sum_{j=0}^{2}\sum_{i=0}^{P}\left(\frac{\mathbf{r}_{yj}^T\mathbf{m}_{ij}}{\mathbf{r}_{zj}^T\mathbf{m}_{ij}} - \frac{Y_i + t_{yj}}{Z_i + t_{zj}}\right)^2 \ , \end{aligned} \qquad (2)$$

where $\mathbf{m}_{ij}$ is the coordinates of the i th feature point of j th image. $\mathbf{r}_{*j}$ is derection column vector of rotation matrix $\mathbf{R}$ of the j th image, t_x, t_y and t_z are elements of translation vector. A 3-D position of i th feature points are represented by X_i, Y_i and Z_i.

We employ Newton method for minimization, which is computational power consuming process if initial value of the parameters are far from real value or there are many parameters for optimization. However, in our approach, high computation performance can be realized because camera poses obtained by motion model are suitably accurate for the initial parameters and the number of parameter can be selected by its accuracy based on the result of motion stereo.

B. Generation of Initial Camera Pose and Object Shape

Initial parameter values for the next *Simultaneous Estimation of Camera and Object Shape* process is generated in this step.

After the *Target Object Detection* process, the robot approaches its target object. To keep good watch for the target object, camera pose controll becomes complex and it causes control delay. This means that there are high possibilities for the camera poses obtained by motion model to include more considerable error. On the other hand, distance between the camera and the target object is more closer, it is possible to acquire more detailed information of camera poses and object shape. Therefore, the revival acquisition of camera pose and 3-D points are performed without motion model.

To re-estimate these parameter values, Factorization method is adopted. A Factorization method can simultaneously calculate camera poses and 3-D position of feature points from a matrix $\mathbf{W}$ which is composed of several feature points in images

$$\mathbf{W} = \mathbf{MS} \ . \tag{3}$$

The matrix $\mathbf{M}$ presents camera poses and the matrix $\mathbf{S}$ presents the 3-D position of feature points. We use Factorization based on weak perspective camera model[7] which can be performed fast but the result includes linear approximation error.

In this step, the robot still approaches to the target object, and it is not so much the case that the feature points of the object are occluded with each other. Then most of feature points can be tracked. Under this condition, factorization method has advantage. After that, the error of Factorization method is compensated by Nonlinear Minimization (Eq.(4)) again, and both accurate camera poses and 3-D position of several feature points are acquired simultaneously. Despite Factorization includes linear approximation error, finally obtained parameter values in this step can be used for good initial values for the next step.

C. Simultaneous Estimation of Camera Pose and Object Shape

Camera poses and object shape are acquired simultaneously and sequentially in this step. As often as new image is obtained, following processes are applied:

A. Camera pose is estimated by means of Nonlinear Minimization from such feature points that are well tracked and their 3-D position is already obtained in the former processes.

B. 3-D position of newly extracted feature points are estimated based on the camera pose which is obtained by A. by means of Motion Stereo.

In this step, the robot begins to move around the object. Feature points will interchange frequently as the viewpoint of the camera changes. In this situation, Motion Stereo is effective because it can calculate the 3-D position of each feature point. Camera pose is estimated from several feature points which have already been calculated, then Nonlinear Minimization is performed to obtain more accurate camera pose.

The function of Nonlinear Minimization here is as follows:

$$\begin{aligned} C = & \sum_{i=0}^{P} \left(\frac{\mathbf{r}_x^T \mathbf{m}_i}{\mathbf{r}_z^T \mathbf{m}_i} - \frac{X_i + t_x}{Z_i + t_z} \right)^2 \\ & + \sum_{i=0}^{P} \left(\frac{\mathbf{r}_y^T \mathbf{m}_i}{\mathbf{r}_z^T \mathbf{m}_i} - \frac{Y_i + t_y}{Z_i + t_z} \right)^2 \ . \end{aligned} \tag{4}$$

In this step, the camera pose is the only parameter to be optimized in Eq.(4). Fast processing can be expected. Motion Stereo is performed to calculate 3-D position of unknown feature points by the estimated camera poses. However there are several feature points which could not obtain accurate 3-D position, each obtained 3-D position of feature points must be reprojected onto the image plane. If the reprojected feature points have good match with the original feature point, it is still accepted as a proper and known feature point.

In this step, each process is fast and reconstruction of the target object can be performed sequentially every time a new image is captured. This enables a robot to plan next camera viewpoint to acquire better shape model from the reconstructed shape in realtime.

V. Improvement of Practical Problems

When proposed algorithm is implemented on a mobile robot, the step of *Simultaneous Estimation of Camera Pose and Object Shape* has several problems as follows:

- Because of mismatched feature correspondence in a series of images by KLT-Tracker, camera pose estimation from feature points whose 3-D position are known is failed.
- Mis-correspondence among images or large 3-D position error of the feature points appears, because the process continues tracking a feature point that should already be hidden by self occlusion.
- If an image is captured when the robot or the camera turns, the image is blurred and it becomes difficult to keep track of many feautre points.

To cope with these problems, our algorithm is improved.

A. Elimination of Mismatched Feature Correspondence

All the captured images from the camera on the robot are not necessarily in good conditions. If an object moves largely in the series of captured images, or if a feature point itself is weak, KLT-Tracker often fails in feature tracking between neighbor two images (Fig.4). Especially, there are a large amount of rotation component to the observed object, the feature tracker produces considerable tracking error.

Estimated camera poses suffer from such ill natured feature points and they become inaccurate. Therefore, we employ the RANSAC to select good feature points according to rules as follows:

- Feature points which have already acquired its 3-D position and are observed between images are selected at random. In our implementation, 60% of feature points are selected from all the points.
- Camera pose is estimated based on the randomly selected feature points. Then we make backprojection of the feature points onto the image plane based on the estimated camera pose and 3-D positions of the feature point. Evaluate squared error between the positions of tracked feature point and back projected feature point in the image.

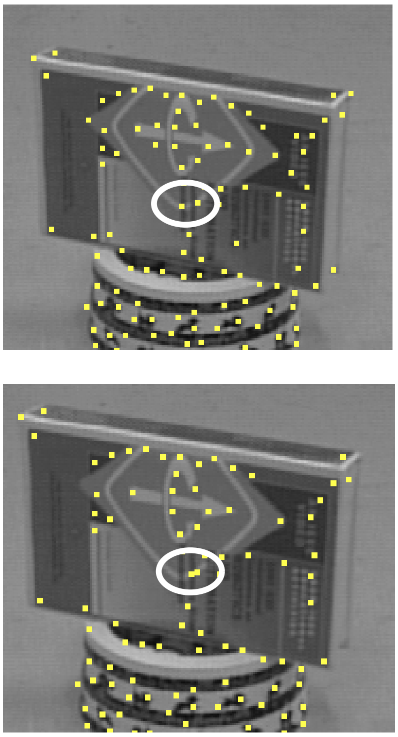
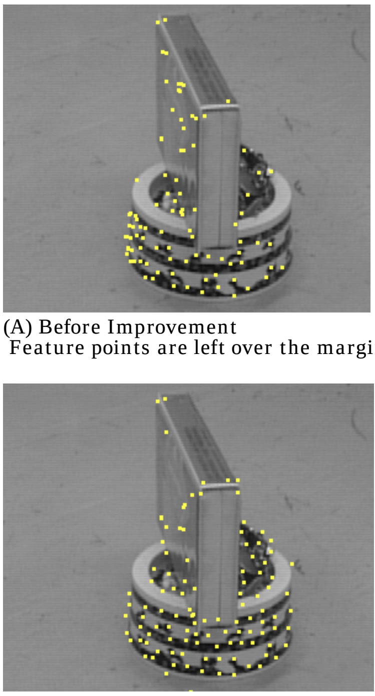


Fig. 4. Failure Tracking



(A) Before Improvement
Feature points are left over the margin

(B) After Improvement
Feature points are extracted equally

Fig. 5. Self Occlusion

Running over this process several hundred times, camera pose can be decided from the best feature point set whose record has highest evaluation.

B. Detection of Self Occlusion

In the third step - *Simultaneous Estimation of Camera Pose and Object Shape* process, appearance or shape of the object in the image changes by degrees as the robot moves around the object. The original feature tracker has a tendency to keep track of a phantom feature even through the feature point is already hidden by self-occlusion.

To cope with this situation, we have improved the original tracker. Feature points in images are checked by investigating relative positions among the noticed feature and the features around it (Fig.5). If the amount of the change in relative positions between adjacent two images exceeds predefined threshold values, feature point is removed. This improvement is effective to remove feature points suffering from tracking error.

C. Detection of Undesirable Captured Image

Fundamentally, all the images which are captured by the camera equipped on the mobile robot are not necessarily in good condition. If the image is captured at the worth condition, blurred image will be obtained. Such an undesirable image gives rise to failure in keeping track of many feature points or increase in mis-track of the feature points. Tracking accuracy will decrease in total. As a result, the process is broken up because camera pose can not be estimated.

A counter measure against this problem is to discard such an undesirable image. Assuming that feature tracking is performed between the $n-1$ th image and n th image, if the number of the feature points that can not be tracked, the n th image is discarded. Then, a new - $n+1$ th image is captured and try the feature tracking between the $n-1$ th and the $n+1$ th image. This image elimination approach provides robust and autonomous process.

VI. Experiments

Experiments are performed with a mobile robot “YAMABICO” which is equipped with a Canon VC-C4 pan-tilt camera. Images are captured by the image processing module HITACHI “IP7500”. Picture resolution is 512×440 pixels and the images are sent to a mobile PC through 10 BASE-T ethernet cable. The images are captured whenever the robot moves 20mm. Feature tracking, estimation of camera pose and object shape are performed in the mobile PC (PentiumM at 1.7 GHz).

For inspecting effectiveness of the proposed shape modeling algorithm, under the condition that several objects are placed on the floor in front of the robot. We issued in advance a command that the robot searches and detects its target object which is in the nearest from the robot. A target object in this experiment has sphere shape which is 220mm in diameter (Fig.6).

Target Object Detection step is performed by 10 images which is captured from start position. Feature points are tracked from these images and 3-D shape is reconstructed from a pair of the first and the 10th images which is captured in this phase. Feature tracking from one image is executed in 450msec, Motion Stereo, Nonlinear Minimization and clustering is executed in 50msec.

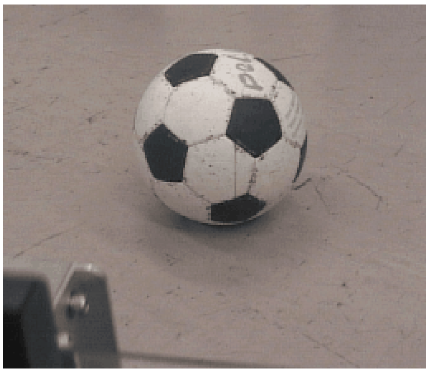


Fig. 6. Original Image

Genaration of Initial Camera Pose and Object Shape step is performed from 25 successive images which are captured from the robot with approaching its target object. The process of acquiring accurate result by Factorization and Nonlinear Minimization is executed in 5sec.

Simultaneous Estimation of Camera Pose and Object Shape step is performed from image streams which are captured around the target object. In this step, because the Motion Stereo needs parallaxes, it is performed between two images whose distance is more than 60mm. Feature tracking including new points is executed in 850msec, and camera pose and 3-D feature points estimation based on RANSAC takes 80msec in each images. In this experiment, as maximum number of execution of RANSAC are fixed 500 counts, the process is finished when the estimated camera pose from Nonlinear Minimization is sufficiently accurate.

From these image streams, 1800 counts of 3-D feature points and 204 camera poses are reconstructed. In terms of accuracy, to the case of real spherical object (Soccer Ball) the error of the shape model is within 10mm. To the case of Glove(Fig.8), height, width and length are 10mm at most, too. These results prove that the result of reconstructed object shape and camera poses has enough accuracy with sufficiently brief processing time for the mobile robot. Fig.8 shows more several instances which are the result of dense shape reconstruction with the method presented [9].

VII. Conclusion

In this paper, we presented fast and robust algorithm to acquire the object shape and the camera poses for a mobile robot. Experimental results are demonstrated and prove the effectiveness of our method.

The robot implements consistent process from searching the target object to reconstructing accurate shape of the object autonomously. Our method is available with simple and convenient sensor system because estimating both of camera poses and object shape are performed from image streams. In the meantime, this method enables viewpoints selection simultaneously because shape reconstruction process is performed sequentialy.

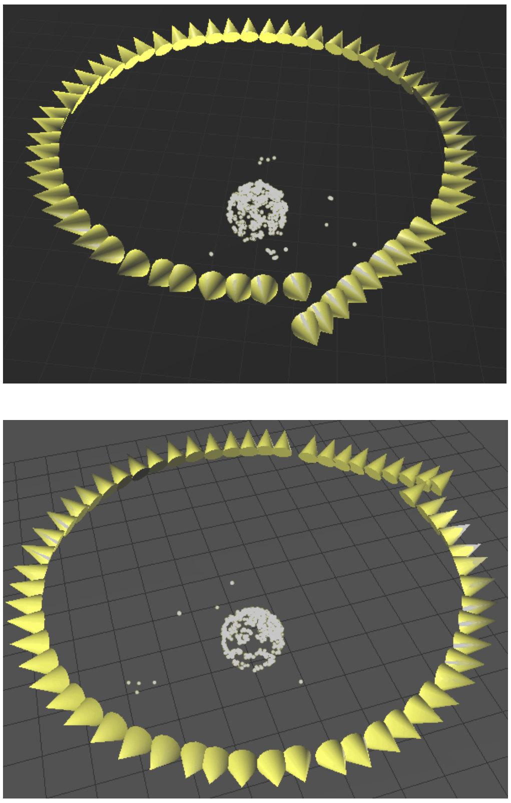


Fig. 7. Reconstructed Camera Pose and Feature Points

As a future work, simultaneous performing both of shape reconstruction and path planning of the robot senser system in real time.

References

[1] C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthgraphy: a Factorization Method", IJCV,9(2), pp.137–154, 1992.

[2] J. Eguchi and S.Yuta: "Grasping and Handing an Object by Autonomous Mobile Manipulator in Office Environment", IFAC Workshop on Mobile Robot Technology, pp.196-201 2001.

[3] J. Shi and C. Tomasi: "Good Features to Track", CVPR, 1994.

[4] J. Miura, Y. Shirai and N. Shimada: "Development of a Personal Service Robot with User-Friendly Interfac es", 4th Int. Conf. on Field and Service Robotics, pp.293– 298, 2003.

[5] K. Nagatani and S. Yuta: "Autonomous Mobile Robot Navigation Including Door Opening Behabior-System Integration of Mobile Manipulator to Adapt Real Environment-", International Conference on Field and Service Robotics (FSR'97) ,,pp.208-215, 1997.

[6] K. Rahardja and A. Kosaka: "Vision-Based Bin-Picking: Recognition and Localization of Multiple Compex Objects Using Simple Visual Cues", in 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems, Osaka, Japan, November", 1996.

[7] Poelman C.J and Kanade T.: "A Paraperspective Factorization Method Shape and Motion Recovery" tech. report CMU–CS–93–219, Computer Science Department, Carnegie Mellon University, 1993.

[8] T. Sato , M. Kanbara, N. Yokoya and H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Handhelds-baseline Stereo using a Hand-held Video Camera", IJCV,Vol 47, No.1-3, 2002.

[9] K.Yamazaki, M.Tomono, T.Tsubouchi and S.Yuta: "3-D Object Modeling by a Camera Equipped on a Mobile Robot" ICRA, 2004.(to appear)

[10] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong.: "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry", Artificial Intelligence Journal, Vol.78, pp.87–119, 1995.

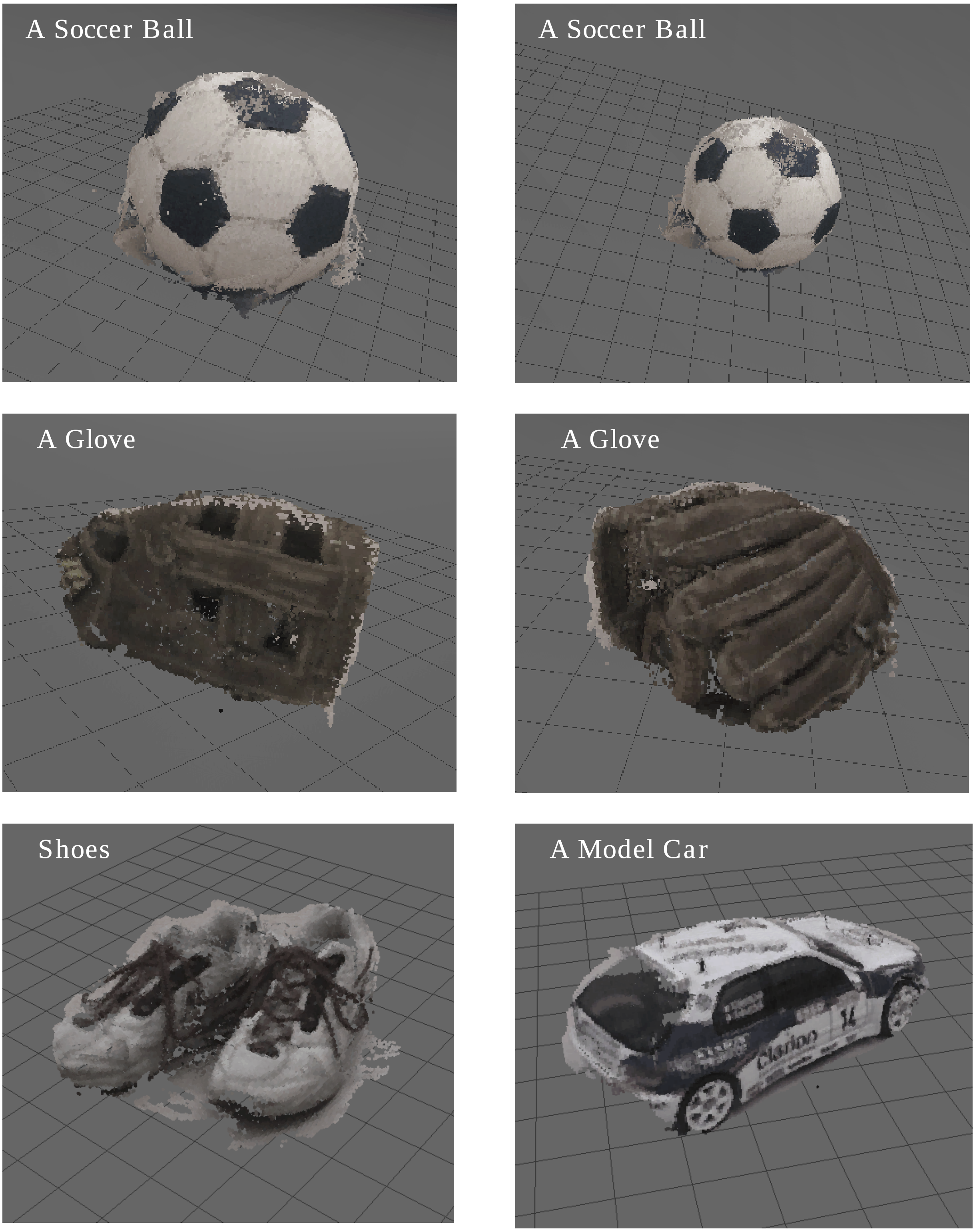


Fig. 8. Reconstructed Dense Shape