The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Real-Time 3D Visual Sensor for Robust Object Recognition

Muhammad Attamimi[1], Akira Mizutani[1], Tomoaki Nakamura[1],
Takayuki Nagai[1], Kotaro Funakoshi[2], and Mikio Nakano[2]
[1]Department of Electronic Engineering, The University of Electro-Communications
1-5-1 Chofugaoka Chofu-shi, Tokyo 182-8585, Japan
[2]Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako-shi, Saitama 351-0188, Japan
{m_att, akira, naka_t, tnagai}@apple.ee.uec.ac.jp, {funakoshi, nakano}@jp.honda-ri.com

Abstract— **This paper presents a novel 3D measurement system, which yields both depth and color information in real time, by calibrating a time-of-flight and two CCD cameras. The problem of occlusions is solved by the proposed fast occluded-pixel detection algorithm. Since the system uses two CCD cameras, missing color information of occluded pixels is covered by one another. We also propose a robust object recognition using the 3D visual sensor. Multiple cues, such as color, texture and 3D (depth) information, are integrated in order to recognize various types of objects under varying lighting conditions. We have implemented the system on our autonomous robot and made the robot do recognition tasks (object learning, detection, and recognition) in various environments. The results revealed that the proposed recognition system provides far better performance than the previous system that is based only on color and texture information.**

I. Introduction

In the area of intelligent systems, the use of 3D information is crucial. The autonomous mobile robots, for instance, are expected to have an ability to recognize their surrounding environments in real time and three-dimensional way in order to complete complex tasks. In recent years the 3D measurement technology has been developed remarkably and many methods have been proposed. These methods roughly fall into two categories. First one is based on passive sensors such as multiple view 3D reconstruction using CCD cameras. Although the baseline stereo camera is widely used these days, quality of the reconstructed 3D information is not high enough in general. Moreover, it is impossible to measure the distance to the textureless region in principle and the computational cost for searching corresponding points is indispensable. On the other hand active sensors including laser range finder (LRF) are very robust to illumination conditions and provide really accurate range information. However, the 3D laser scanning system is large and very expensive. Hence, such system is not suitable for home assistant mobile robots. Instead of using expensive laser scanner system, it is possible to use the motor-driven horizontal line laser scanner for 3D SLAM in mobile robot applications. Such system, however, does not provide 3D information in real time.

The infrared ray (IR) time-of-flight (TOF) cameras [1], which overcome these problems, are of current interest, since they are capable of obtaining 3D information around 30 fps. The TOF camera is relatively small and less expensive compared with 3D LRF. Although the TOF camera is not comparable in accuracy to 3D LRF, its performance is far better than that of commercially available stereo camera. These active measurement systems still have the problem that no color information at each pixel location is accompanied.

In this paper we present a realization of novel 3D measurement system, which yields both depth and color information in real time by integrating a TOF and two CCD cameras. The calibration is carried out by estimating the internal and external camera parameters for both TOF and CCD cameras. Another problem to be solved in this paper is occlusions between TOF and CCD cameras. These occlusions occur because the viewpoints of all sensors are different. To cope with this problem fast occluding pixel detection algorithm is proposed. Since the system consists of two CCD cameras, missing color information of an occluded pixel is compensated by the other camera. The proposed 3D visual sensor makes it possible to simultaneously capture accurate 3D information and corresponding color information in real time.

The latter half of this paper is devoted to the proposed object recognition system using the 3D visual sensor. Thanks to the real-time 3D visual sensor, the integration of multiple cues, such as color, texture and depth, is possible for object recognition. The depth information, among others, improves robustness considerably. As a feature vector for the depth information, Histogram of Depth (HOD) is proposed in this paper. We also combine Shape Distribution (SD) [2], which is normally utilized for the 3D model retrieval, with HOD. These features are invariant to scale and rotation, which is a desirable property for object recognition. Furthermore, HOD can be computed really fast. As for color and texture, histograms of hue and vector quantized dense SIFT (Scale Invariant Feature Transform) are utilized, respectively.

Related works include applications of TOF cameras and a huge number of works for object recognition. As we mentioned earlier, IR-TOF camera is of current interest and some applications have been proposed for the last few years. In [3], 3D SLAM using TOF camera for rescue robots has been proposed. The estimation of camera motion and 3D pose has been proposed in [4] and [5]. The TOF camera has been also found application in head pose and object tracking [6][7]. These works, however, use only 3D information which is captured by TOF camera. On the one hand, [8] has proposed the combination of TOF and CCD cameras for virtual reality applications. However, the work does not deal with problems of the calibration and occlusions between sensors. There are very few examples of the home assistant robot with calibrated TOF and CCD cameras at present. Although the range calibration for TOF camera has been discussed in [9][10], few papers mention the compensation of missing color information caused by occlusions between sensors.

Regarding object recognition, existing methods are

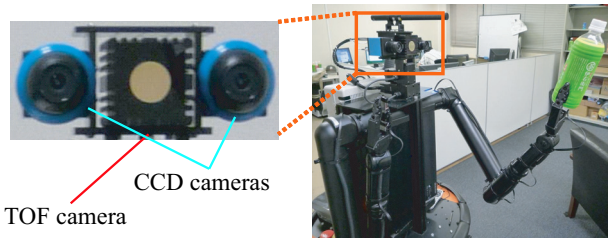

Fig. 1. DiGORO, the robot with the proposed 3D visual sensor.

roughly divided into two groups; appearance (2D)-based method and 3D shape-based one. The matching of local features, e.g. SIFT [11], is now the most popular method of appearance-based object recognition. It is well known fact that SIFT is invariant to scale change and rotation. However, it cannot deal with affine transform. In [12], Bag of Keypoints model, which utilizes the histogram of vector quantized SIFT(visual words), has been discussed. In either case, it is difficult to recognize textureless objects. In the area of 3D model retrieval, Shape Distribution [2] has been proposed for measuring the similarity between two 3D models. Although the SD works well for the 3D model retrieval, it is obvious that discrimination of different objects, which have similar shapes with different colors and/or textures, is impossible. To be precise, there is a difference between 3D model retrieval and the problem considered in this paper, i.e. object recognition based on 2.5 dimensional depth information.

In [13], the original 2D SIFT has been extended to N-D SIFT. The N-D SIFT has been applied to MRI data and proven to be beneficial in matching of N-D volume data. Since the proposed visual sensor gives 2.5D information, the application of N-D SIFT is not straightforward. Moreover, N-D SIFT is not suitable for real-time robotic applications, since the computational cost of N-D SIFT is too high because of its very high dimensional feature vector.

In [14], authors have proposed color CHLAC features, which are based on 3D and color information, for detecting the object in a complex 3D scene. However, the color CHLAC feature is not rotation invariant. Moreover, it seems that the feature is affected to some degree by changes in illumination since color information is directly incorporated in the feature.

II. Realization of 3D Visual Sensor

The proposed vision sensor is depicted in Fig.1. The sensor can acquire color and accurate depth information in real time by calibrating a TOF and two CCD cameras. In order to calibrate two different types of cameras, following problems are required to be considered. First, the CCD camera has much higher resolution (1024×768) than the TOF camera (176×144). Second, each camera has its own parameters such as focal length, lens distortion, relative position, and so forth. The occlusion problem, which caused by relative positions of sensors, has to be resolved as well. These problems are discussed in this section.

A. TOF Camera

The distance measurement capability of TOF camera is based on the TOF principle. In TOF systems, the time taken for light to travel from an active illumination source to the objects in the field of view and return to the sensor is measured. In this paper, an off-the-shelf TOF camera SwissRanger SR4000 [15] is used. It emits a modulated near-infrared (NIR) and the CMOS/CCD imaging sensor measures the phase delay of the returned modulated signal at each pixel. These measurements in the sensor results in a 176×144 pixel depth map.

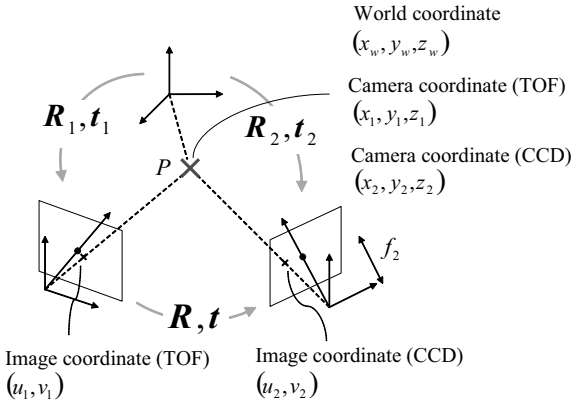

Fig. 2. Coordinate system transformation.

(a) (b) (c)

Fig. 3. An example of the color mapping: (a)color image (1024×768), (b)depth image (176×144), and (c)mapped color image (176×144).

B. Camera Calibration

In the geometric camera calibration, the parameters that express camera pose and properties can be classified into extrinsic parameters (i.e. rotation and translation) and intrinsic ones (i.e. focal length, coefficient of lens distortion, optical center and pixel size). The extrinsic parameters represent camera position and pose in 3D space, while the intrinsic parameters are needed to project a 3D scene onto the 2D image plane.

We use Zhang's calibration method in our proposed system, since the technique only requires the camera to observe a checkerboard pattern shown at a few different orientations. For the calibration of TOF camera, the reflected signal amplitude can be used to observe the checkerboard pattern. Therefore, it is straightforward to apply the same calibration method.

C. Color Mapping

Suppose that the geometric relationship between the camera coordinate systems is represented as in Fig. 2. Then, the coordinate of a point P can be written as,

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = [\mathbf{R}|\mathbf{t}] \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix}, \quad (1)$$

$$\mathbf{R} = \mathbf{R_2}\mathbf{R_1^{-1}}, \qquad \mathbf{t} = \mathbf{t}_2 - \mathbf{R_2}\mathbf{R_1}^{-1}\mathbf{t}_1, \quad (2)$$

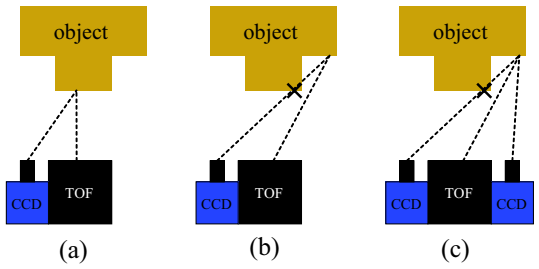

Fig. 4. An illustration of occlusion: (a) the point that can be measured by all sensors, (b) the point that cannot be measured by some sensor, and (c) compensating data by other sensors.

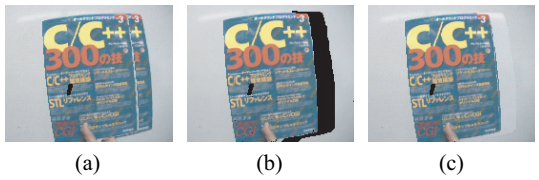

Fig. 5. Detection and compensation of occluded regions: (a)color mapping without removal of the occluded regions, (b)detection of occluded regions, and (c)compensation of occluded regions.

where $(\mathbf{R}_1,\ \mathbf{t}_1)$ and $(\mathbf{R}_2,\ \mathbf{t}_2)$ denote extrinsic parameters for TOF and CCD cameras, respectively. $\mathbf{R}$ and $\mathbf{t}$ represent the rotation matrix and the translation vector for the CCD camera. It should be noted that lens distortions are assumed to be compensated by the estimated intrinsic parameters. Since the TOF camera provides (x_1, y_1, z_1) directly, we can transform 3D information for all pixels of the TOF camera into the CCD camera coordinate system using (1). Finally, color mapping is carried out by the following perspective projection,

$$u_2 = f_2 \frac{x_2}{z_2}, \qquad v_2 = f_2 \frac{y_2}{z_2}, \tag{3}$$

where (u_2, v_2) and f_2 represent pixel location on the CCD's image plane and the focal length of the CCD camera, respectively. Fig. 3 shows a result of the color mapping process.

D. Occlusion

Occluded regions must be considered in the integration of multiple sensors. When two sensors are involved, the following four regions are possible:

1) The region which is observable from both of TOF and CCD cameras (Fig. 4(a)).
2) The region which is observable only from the TOF camera (Fig.4(b)).
3) The region which is observable only from the CCD camera.
4) The region which is unobservable from both of TOF and CCD cameras.

Since color information, which is captured by the CCD camera, is mapped to a corresponding pixels of the TOF's depth map, we have to take care of the second case. In this case, the CCD camera observes a different point (marked with $\times$ in Fig.4(b)) from the point where the TOF camera sees. Obviously this situation leads to a false mapping of color information and a pseudo object appears as shown in Fig.5(a). Therefore, occluded regions have to be detected and removed. The Z-buffer method, which is widely used in the area of computer graphics, is applicable. However, it requires a 3D data with polygon mesh representation. Instead, we propose a fast occlusion detection algorithm that works in the color mapping process. The basic idea is to use the property for alignment sequence of corresponding pixels between TOF and CCD cameras. If there is no occlusion, the corresponding pixel location of the CCD camera moves to the same direction as the horizontal scanning direction of the TOF camera's depth map. On the other hand, the corresponding pixel location of the CCD camera moves to the opposite direction when the occluded region starts. The algorithm for detecting occluded regions is described as following steps:

1) Find all pixels in the CCD camera that have correspondence with $N_u \times N_v$ pixels of the TOF camera. Now, let $(U(u_2), V(v_2))$ be a coordinate of the CCD camera's image plane that corresponds to (u_2, v_2) on the TOF camera. Set $u_2 = 2$ and $v_2 = 1$.
2) Find the pixel which can be written as $U(u_2) < U(u_2 - 1)$, for any v_2. If it is found then go to 3), if it is not found increase u_2 and keep searching. If $u_2 = N_u$ go to 4).
3) Find the pixel which can be written as $U(u_2) < U(u_2 + k)$ by increasing k. If it is found, set $u_2 = u_2 + k$ then go back to 2), if it is not found then remove corresponding point of $(u_2 + k, v_2)$ and go back to 3). If $u_2 + k = N_u$ then go to 4).
4) Increase v_2 and set $u_2 = 2$, then go back to 2). The process is finished when $v_2 > N_v$.

Fig.5(b) shows a result of the proposed occlusion detection. It should be noted that the algorithm above is for the left-CCD and TOF cameras. For the right-CCD camera, the scanning has to carried out from right to left direction (from $u_2 = N_u$ in decreasing direction).

Since our proposed system has two CCD cameras, which are mounted on both sides of the TOF camera, color information of occluded pixels are compensated by the other camera as shown in Fig.4(c). More precisely, the left-CCD camera is mainly used for the color mapping in our system and the right-CCD camera is used for compensating occluded pixels.

III. Robust Object Recognition

In this section we discuss a robust object recognition method using the 3D visual sensor described in the previous section. The overview of the proposed method is illustrated in Fig.6.

A. Object Detection

Object detection is required as the first step in the object learning and object identification subsystems. In general, object detection under complex background is not an easy task. Here we use the following two different methods, which use the information acquired by the proposed sensor.

1) Object Detection based on Motion Attention: When the user shows a target object to the robot, the object can be segmented out by paying attention to the motion cue. The motion-attention based method uses a motion detector for extracting an initial object region and then the object region is refined using the color and depth information of the initial region. Please refer to [16] for more details. An example of object detection based on the motion attention is shown in Fig.7.

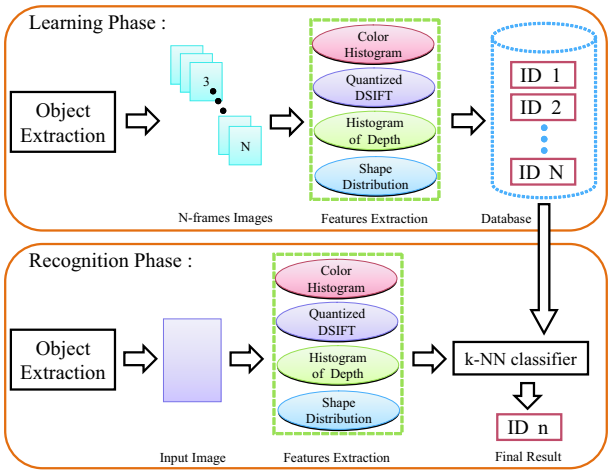

Fig. 6. Schematic of the proposed recognition system.

(a) (b) (c)

Fig. 7. The result of object detection; (a)input image, (b)object probability map, and (c)extracted object.

2) Object Detection based on Plane Detection: If objects are on the table, the plane detection technique is beneficial to detect the objects. The 3D Randomized Hough Transform (3DRHT) is utilized for fast and accurate plane detection. Since we can acquire highly accurate 3D information from the TOF camera, we can detect plane in high accuracy. Then, all 3D points except for top of the table itself are projected onto the detected plane. Finally, connected components analysis (labeling) is performed on the plane and each object is segmented out. It is worth noting that high resolution image for the segmented object can be obtained by using the correspondence between TOF and CCD cameras. Fig.8 shows the result of detected objects on the table.

B. Object Learning

Since k-Nearest Neighbor (k-NN) classifier is used for the object identification, object database is generated in the learning phase. The user is supposed to show the target objects in various directions to the robot in the learning phase. The other scenario is that the robot detects the object on the table and observes it in various directions by grasping it. In any case, object detection is carried out in many frames and feature vectors are computed. These feature vectors are stored for the object database. Here, we need features which are invariant to scale change, rotation, and translation. The histogram-based feature is employed on our proposed method because it makes features rotation and translation invariant. Moreover, we can use 3D information for normalizing the scale, which results in the scale invariant features. Since it is difficult to realize the feature that has view invariance, this property is ensured by matching of all features from various view points, which are accumulated

(a) (b)

Fig. 8. Object detection based on plane detection. (a)Detected plane, and (b)Detected objects.

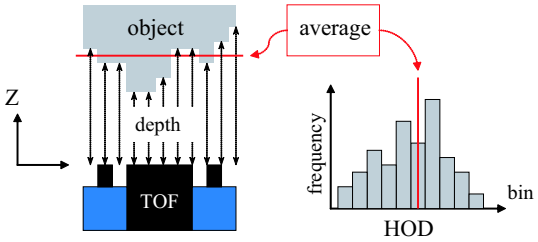

Fig. 9. Schematic of HOD.

in the learning phase. We will explain the features that are employed on this research.

1) Histogram of Depth: As one of features using 3D information, we propose Histogram of Depth (HOD). This is a histogram of depth values of all pixels in an object region as shown in Fig.9. Since the histogram would vary by the distance to an object, we normalize the histogram using the mean distance of the object region. In consideration of the accuracy of the TOF camera, the histogram bin size is chosen to be 5 mm. The HOD is very fast computable and is also invariant to scale change, rotation, and translation. However, it should be combined with another feature, since it is not powerful enough to distinguish the shapes of objects.

2) Shape Distribution: We use Shape Distribution (SD) as the second feature. The SD represents various characteristics of the object 3D shape by calculating a metric among vertices. We utilize distances between all combinations of two vertices in the object region as shown in Fig.10, followed by taking the histogram of these distances as an object feature. The SD has been proposed to measure the similarity between 3D models. Although the SD for 3D models has view invariance property, it is viewpoint dependent in our case since the depth information is 2.5D in principle. There is no need for generating point clouds randomly from 3D polygon mesh, since 3D information is obtained as point clouds in our proposed system. Moreover, the SD is invariant to scale change, rotation, and translation. However, it has high computational complexity since all combinations of two vertices in the object is taken into consideration. We solve this problem by decimating vertices. In fact, the decimation by a factor of five makes the computation reasonably fast without degrading the recognition performance. By combining HOD and SD, improvement of the recognition performance can be expected since SD contains size information of the object.

3) Color and Texture Information: We have discussed features, which are based on 3D information. In order to distinguish similar shaped objects with different color and/or textures, we also utilize color and texture information for the object identification. Here, the 2D histogram of

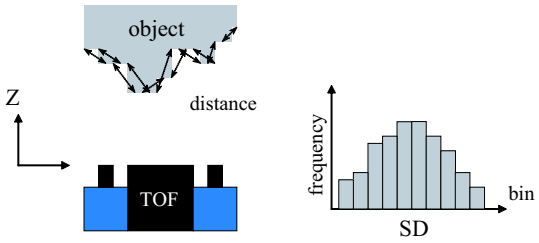

Fig. 10. Overview of SD.

(a) (b)

Fig. 11. Experimental environment: (a)living room, and (b)objects used in the experiment.

place1 (approx. 400Lux) place2 (approx. 650Lux) place3 (approx. 150Lux) place4 (approx. 50Lux)

Fig. 12. Varying illumination conditions.

category1 category2 category3 category4 category5

Fig. 13. Object categories.

hue and saturation in HSV color space is used as color information. As for texture, we utilize a dense version of SIFT (DSIFT). Unlike the original SIFT, DSIFT is not invariant to scale change since it omits the time-consuming key points detection and calculates densely sampled features. Instead, the scale change can be normalized using the depth information in our system. Before taking histogram, DSIFT is vector quantized using the predetermined 500 dimensional codebook, which is generated by k-means clustering of many images of indoor random scenes.

C. Object Identification

In the object identification subsystem, the object region is segmented out at first using the motion-attention based method or the plane detection based one. Then, the feature vector is extracted from the input object region. The identification is based on Bhattacharyya distance between the input feature vector and the reference one, which is selected from the database. Now let the histogram of a reference data, the histogram of an incoming data, and size of the histogram be H_{ref}, H_{in}, and N, respectively. Bhattacharyya distance $D(H_{ref}, H_{in})$ can be written as

$$D(H_{ref}, H_{in}) = \sqrt{1 - \sum_{i=1}^{N} \sqrt{H_{ref}(i) \times H_{in}(i)}}. \quad (4)$$

Likelihood P is given by

$$P = \exp\{-\lambda\{D(H_{ref}, H_{in})\}^2\}, \quad (5)$$

where λ is a coefficient to adjust the variance of D and dependent on the type of a feature. This coefficient is set to be inversely proportional to the average of D^2 for specific feature vectors in the learning data and works as a weight for each feature in the recognition phase. The likelihood P is calculated for all features, i.e. color, texture, and depth, and then the integrated likelihood for the candidate object is obtained by multiplying all of these likelihood values. Furthermore, reliability of color and texture information largely depends on the illumination condition during the recognition. Here, we presume that the maximum likelihood value of all likelihoods, which are calculated for all data in the database, as a reliability value of that feature. We do not use the feature if the maximum likelihood value is lower than the predefined threshold.

Finally, k-NN classifier is employed to make a final decision. Since we are aiming at robust object recognition using a robot in this research, it is also possible to change the viewpoint actively and observe an object in multi-frame, if the reliability of the current input is low. In particular, more distinctive 3D information could be obtained if the robot changes the viewpoint. Thus, it is possible to improve the recognition performance by integrating multi-frame information.

IV. Experiment

An experiment is carried out in the living room shown in Fig.11(a) using 20 objects in Fig.11(b). A user shows each object in various angles to the robot and database, which contains feature vectors of 50 frames per object, is generated by the robot in the learning phase. In the recognition phase, 200 frames of images for each object are identified. These images are captured at four different locations (50 frames per location), which have different illumination conditions as shown in Fig.12. It should be noted that the learning phase is carried out in the location 1. We use three different combinations of features as color and texture (2D information), HOD and SD (3D information), and integration of all features. The standard SIFT approach is also involved for the purpose of reference. Since the recognition performance depends deeply on characteristics of the target objects, we categorize 20 objects into 5 categories as shown in Fig.13 and the recognition rate is computed for each category.

Fig.14 shows the results of recognition for each object category. The results for each location are given in Fig.15.

A. Recognition by Color and Texture Information

It can be seen from Fig.14 that objects in categories 1 to 3, which are objects with well-defined color and/or have many textures, can be recognized almost perfectly using color and texture information. As a matter of course, it is difficult to recognize white objects in categories 4 and 5 under varying illumination conditions.

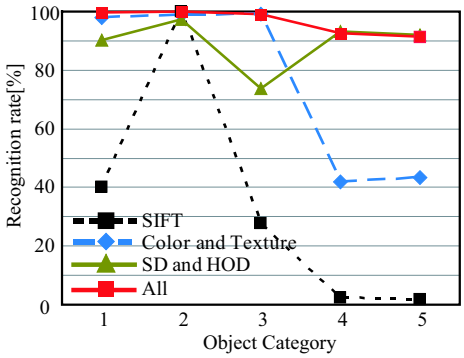

Fig. 14. Recognition results for each category.

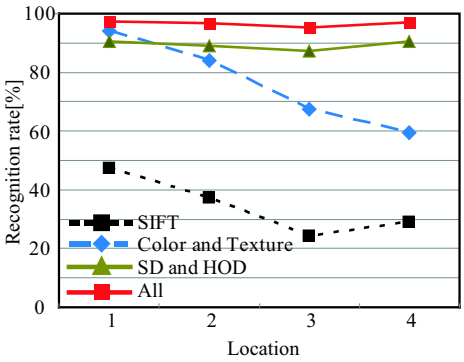

Fig. 15. Recognition results for each location.

The standard SIFT approach shows a similar tendency to this case. The color information gives rise to difference in the recognition rate since color information is not included in SIFT.

B. Recognition by 3D Information

By using 3D information, fairly different behavior of the recognition result from the one using 2D appearance is observed. First of all, it can be confirmed from Fig.15 that HOD and SD are not sensitive to changes in illumination. Moreover, Fig.14 suggests that HOD and SD work reasonably well for white objects in categories 4 and 5, which are very difficult to recognize using color and/or texture information.

On the contrary, performance degradation is observed for the objects in categories 1 to 3. The reason for this degradation is the similarity in shape. Particularly with the category 3, objects are similar in shape and size. In spite of the common flat surface, the objects in category 2 are discriminable because of the difference in size.

C. Recognition by Integrated Information

Not surprisingly, the integrated features yield the best result. It can be seen that the recognition rate for pure white objects with no texture (e.g. white dish, mug etc.) is as high as 90%. Furthermore, it can be easily confirmed that the proposed recognition system is extremely robust to the illumination change.

It seems that the change of viewpoint to an object mainly caused the false recognition in this experiment. It is impossible to observe an object from equally spaced viewpoints, since a user shows the object to the robot. These kind of biased samples are responsible for deviation of the voting process.

V. Conclusion

In this paper, we have proposed 3D visual sensor based on the calibration of TOF and two CCD cameras. The proposed sensor can provide accurate 3D depth with registered color information in real time. We have also proposed a robust object recognition system by integrating color, texture, and 3D information based on the proposed sensor. The results given in this paper indicates that the proposed system is very robust to illumination change and applicable to various objects.

Currently, we are working on the development of a robot that can learn objects in a perfectly automatic manner based on the proposed method. Improvement of the recognition performance is also an issue in the future.

References

[1] T.Oggier, F.Lustenberger and N.Blanc, “Miniature 3D TOF Camera for Real-Time Imaging”, in Proc. of Perception and Interactive Technologies 2006, pp. 212-216, June2006
[2] R. Osada, T. Funkhouser, Bernard Chazelle, and David Dobkin, “Shape Distributions”, ACM Transactions on Graphics, vol.21, no.4, pp.807-832, Oct.2002
[3] K.Ohno, T.Nomura and S.Tadokoro, “Real-Time Robot Trajectory Estimation and 3D Map Construction using 3D Camera”, in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.5279-5285, Oct.2006
[4] S.May, D.Droeschel, D.Holz, C.Wiesen and S.Fuchs, “3D Pose Estimation and Mapping with Time-of-Flight Cameras”, in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Oct.2008
[5] C.Beder, I.Schiller and R.Koch, “Real-Time Estimation of the Camera Path from a Sequence of Intrinsically Calibrated PMD Depth Images”, in Proc. of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol.XXXVII, pp.45-50, July2008
[6] S.B.Gokturk and C.Tomasi, “3D Head Tracking Based on Recognition and Interpolation Using a Time-Of-Flight Depth Sensor”, in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.211-217, July2004
[7] D.W.Hansen, M.S.Hansen, M.Kirschmeyer, R.Larsen and D.Silvestre, “Cluster tracking with Time-of-Flight cameras”, in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp.1-6, June2008.
[8] B.Bartczak, I.Schiller, C.Beder and R.Koch, “Integration of a Time-of-Flight Camera into a Mixed Reality System for Handling Dynamic Scenes, Moving Viewpoints and Occlusions in Real-Time”, in Proc. of International Symposium on 3D Data Processing, Visualization and Transmission, June2008
[9] S.Fuchs and G.Hirzinger, “Extrinsic and depth calibration of TOF-cameras”, in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp.1-6, June 2008
[10] Y.Min Kim, D.Chan, C.Theobalt and S.Thrun, “Design and Calibration of a Multi-view TOF Sensor Fusion System”, in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp.1-7, June2008
[11] D.G.Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, Vol.60, No.2, Nov.2004
[12] G.Csurka, C.Dance, L.Fan, J.Willamowski, and C.Bray, “Visual Categorization with Bags of Keypoints”, Int. Workshop on Statistical Learning in Computer Vision, pp. 1-22, May2004
[13] W.Cheung, and G.Hamarneh, “N-SIFT : N-Dimentional Scale Invariant Feature Transform for Matching medical Images”, in Proc. of IEEE Int. Symposium on Biomedical Imaging, pp.720-723, Apr.2007
[14] T.Harada, A.Kanezaki, Y.Kuniyoshi, “The Development of Color CHLAC Features for Object Exploration in 3D Map”, Journal of Robotics Society of Japan, vol.27, no.7, pp.749-758, Sep.2009(in Japanese)
[15] MESA Imaging, http://www.mesa-imaging.ch/index.php
[16] M.Attamimi, A.Mizutani, T.Nakamura, K.Sugiura, T.Nagai, N.Iwahashi, H.Okada and T.Omori, “Learning Novel Objects Using Out-of-Vocabulary Word Segmentation and Object Extraction for Home Assistant Robots”, in Proc. of Int. Conf. on Robotics and Automation, pp. 745-750, May2010