In: Proceedings of 15th Towards Autonomous Robots, TAROS 2014, LNAI 8717, pp. 273-274, Springer Heidelberg. Birmingham, UK, October 1-3, 2014.

Learning Objects From RGB-D Sensors for Cleaning Tasks Using a Team of Cooperative Humanoid Robots

Marcelo Borghetti Soares, Pablo Barros, and Stefan Wermter

University of Hamburg, Department of Computer Science,
Vogt-Koelln-Strasse, 30, 22527, Hamburg, Germany
{borghetti,barros,wermter}@informatik.uni-hamburg.de
http://www.informatik.uni-hamburg.de/WTM

In this work, we address the problem of implementing cooperative search in humanoid robots (NAOs). The robots are taught to recognise a number of objects and then use their RGB-D sensors (attached to their heads) to search their environment for these objects. When an object is found they have to move to the target position and to perform a cleaning task in these objects (also, the location of recognized objects can help navigation). The challenge is threefold: 1) navigation/exploration, 2) real-time object recognition and 3) cooperation. This work will show preliminary results in object recognition and briefly discuss the approaches that will be employed for the entire system.

The scenario consists of a room in which some objects are spread (Figure 1 on the left). Initially, the robots (Figure 1 on the right) explore the environment until their RGB-D devices have detected objects. Then, the robots move to the objects, constantly trying to identify them. It is an assumption of our approach that these objects are located on tables or plane surfaces, so this strategy will help to detect potential objects before recognizing them. The robot has to extract the appropriate information from the point cloud, filter noise and correctly segment the objects. We are using RANSAC (RAndom SAmple Consensus) [1] to identify planes and VFH (View Point Feature Histogram) to collect a multidimensional descriptor (feature vector with 308 elements) that characterizes the object. The robots are, initially, completely unaware about the object's position. Thus, they cannot plan the best way to cooperatively distribute themselves in the environment, but they can model the influence of other robots as repulsive potential fields, in a purely reactive, though collaborative, way. Although a navigation approach based on RGB-D data is intended to be used, initially we will employ a ceiling camera (also used in [2]) that provides a global view of the environment.

Figure 2 shows some objects used in our experiments. We decided to create our own database to employ the same objects in a future stage of the project for online recognition. We have collected RGB-D data of 5 different types of objects (category) and 5 different objects of each type (instances) in 12 different views. The total number of images acquired was 300. We also applied rotation (roll, pitch and yaw) and added noise (in 10% of the points of each point cloud) to create a dataset of 4200 samples. From this total, 75% were used for training and 25% used for testing. A Multichannel Convolutional Neural Network from

In: Proceedings of 15th Towards Autonomous Robots, TAROS 2014, LNAI 8717, pp. 273-274, Springer Heidelberg. Birmingham, UK, October 1-3, 2014.

Fig. 1. (Left) Final scenario to be used and (Right) NAO robot with RGB-D device.

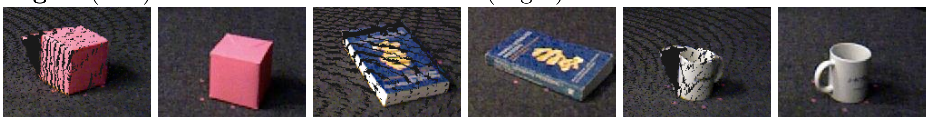

Fig. 2. Examples of RGB-D of 3 objects: box, book, cup.

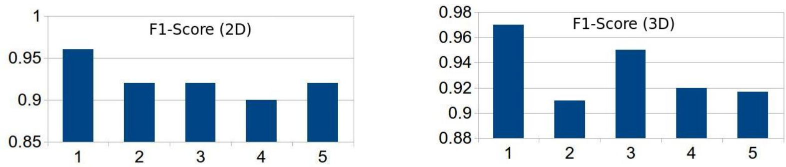

Fig. 3. *F1-score* results for Book(1), Box(2), Can(3), Cup(4) and Sponge(5)

our previous work [3], developed for gesture recognition, was used and trained with the *Backpropagation Algorithm*. The *F1-score* results for $2D$ images and $3D$ point clouds can be seen in Figure 3. These results show that our approach is promising, with *F1-scores* values greater than 0.9 for all objects. The next challenges that should be addressed are: 1) how to combine efficiently $2D$ and $3D$ information, 2) how to embed this recognition system in the robots to perform real time recognition and 3) treat noise and the presence of multiple objects. We expect that in the future the benefits of employing multiple robots for assistance living surpass the costs expended to acquire the technology.

References

1. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24** (1981) 381–395
2. Yan, W., Weber, C., Wermter, S.: A hybrid probabilistic neural model for person tracking based on a ceiling-mounted camera. JAISE **3** (2011) 237–252
3. Barros, P., Magg, S., Weber, C., Wermter, S.: A Multichannel Convolutional Neural Network for Hand Posture Recognition. In: International Conference on Artificial Neural Networks, Accepted (2014)