

RGB-D-Based Features for Recognition of Textureless Objects

Santosh Thoduka, Stepan Pazeckha, Alexander Moriarty, and
Gerhard K. Kraetzschmar

Bonn-Rhein-Sieg University of Applied Sciences
Department of Computer Science
Grantham-Allee 20, 53757 Sankt Augustin, Germany
`<first_name>.<last_name>@inf.h-brs.de`

Abstract. Autonomous industrial robots need to recognize objects robustly in cluttered environments. The use of RGB-D cameras has progressed research in 3D object recognition, but it is still a challenge for textureless objects. We propose a set of features, including the bounding box, mean circle fit and radial density distribution, that describe the size, shape and colour of objects. The features are extracted from point clouds of a set of objects and used to train an SVM classifier. Various combinations of the proposed features are tested to determine their influence on the recognition rate. Medium-sized objects are recognized with high accuracy whereas small objects have a lower recognition rate. The minimum range and resolution of the cameras are still an issue but are expected to improve as the technology improves.

Keywords: object recognition, machine learning, textureless objects, RGB-D data, coloured pointclouds

1 Introduction

As industrial robots become increasingly autonomous there is a need for sophisticated perception capabilities. In controlled industrial settings where the environment is well described, perception tasks are simplified since assumptions can be made about the location of objects. As a result of assuming the object location is known, object recognition may not be required and object detection may be sufficient. However, with mobile robots, the same simplifications cannot be made since there is more uncertainty about the environment. Although the general locations of objects are known, the robot can no longer rely on being precisely localized in the environment. To compensate for this, there is a greater emphasis on performing complex perception tasks such as object recognition.

The availability of low cost RGB-D cameras has progressed research in 3D object recognition significantly. However, industrial objects pose a challenge for existing object recognition approaches due to their nature. Objects such as profiles, nuts, screws and bolts tend to be textureless, of homogeneous colour and, in some cases, quite small. Many are simple geometric shapes made of metal or

plastic and often do not have particularly distinguishable features. There are also similarly shaped objects with only size or colour differentiating them. Existing state of the art 3D recognition algorithms rely on having sufficiently detailed point clouds of objects in order to extract features such as surface normals, colour gradients etc. For small objects, this is a challenge due to the minimum range of RGB-D cameras and their resolution.

RoboCup@Work [9] and RoCKIn@Work [4] are both robotic competitions which focus on mobile manipulation challenges relevant for small and medium sized industrial factory settings. In larger, traditional factories, machinery, service areas and robots can be fixed for long term production where the factory layout and production process is not expected to change frequently. In small factories settings, specifically Factories of the Future [1] which can adapt quickly and dynamically to meet production demands, a particular service area may serve multiple purposes through the production process. Service areas are locations where manipulation and perception tasks are performed. They are general purpose areas which may be shared with humans. As such, service areas can be cluttered and the location of objects on them not precisely known.

In RoboCup@Work, several tasks involve grasping objects that are placed on service areas among other objects. The objects need to be recognized and transported to different locations based on the task specification. In some cases, objects need to be inserted into containers or cavities. Some examples of the industrial objects used in the competition can be seen in Figure 1.

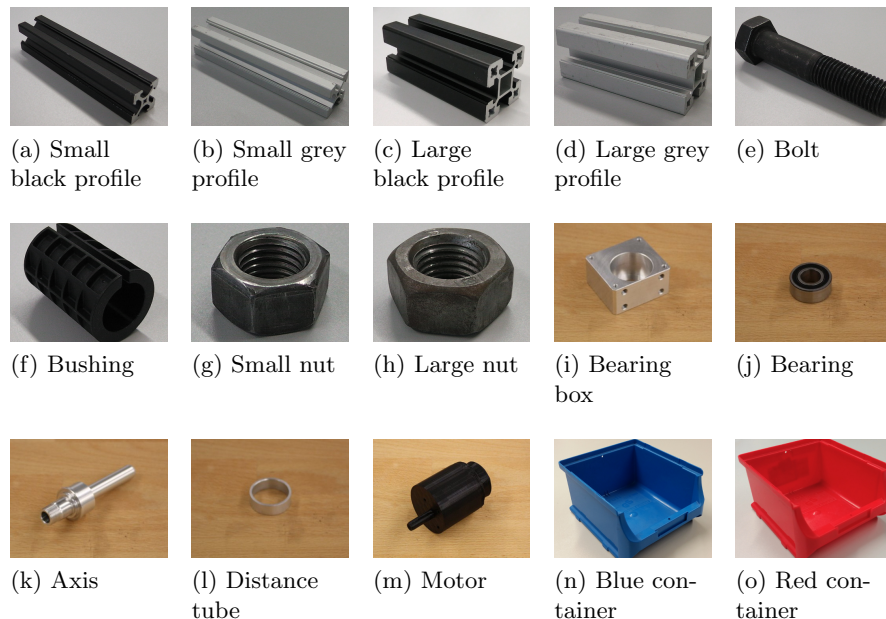


Fig. 1: Object set used in the RoboCup@Work competition [3]

Currently, this is the exact set of objects used, and there are no variations among each type.

The task of object recognition usually involves an offline training phase and an online recognition phase. In the training phase, representative samples of the objects are collected. For 3D recognition systems, this is typically point clouds of the objects taken from several views. Descriptive features are then extracted from the samples and used to train a classifier or save templates. During the recognition phase, an unknown object is segmented from the scene and the identical features are extracted from it. The features are then fed into the classifier or template matcher which returns the identifier of the best matched object from the previously trained objects.

As seen in Figure 2, the point clouds generated by the RGB-D camera are noisy and does not capture all the small details of the objects. The small size and inadequately descriptive point clouds make the task of recognizing such objects a challenging one. For example, the large aluminium profile is only 10 cm x 4 cm x 4 cm, and the distance tube is 1 cm high with a radius of 1.6 cm. They are quite small in the field of view of the camera and the number of points that represent the object is quite low in some cases. In this paper, we focus on the extraction of descriptive features for textureless objects and test the approach using the objects in Figure 1.

The paper is structured as follows: We review related work in Section 2, describe our approach in Section 3 and finally present the results in Section 4.

2 Related Work

Object recognition using 3D point clouds can be broadly categorized into global and local feature-based methods. Global feature descriptors are computed for the entire object point cloud whereas local descriptors are calculated for individual points in the cloud. For example, the Point cloud library (PCL)[12] has implementations for local descriptors such as Point Feature Histogram (PFH), Radius-based Surface Descriptor (RSD), Signatures of Histograms of Orientations (SHOT) and global descriptors such as Viewpoint Feature Histogram (VFH), Ensemble of Shape Functions (ESF) and Global Fast Point Feature Histogram (GFPFH). These descriptors calculate relationships between points such as distances, angles of surface normals etc. and build histograms to represent the distribution of these relationships for each object. During the recognition phase, the stored descriptors are compared with descriptors calculated on the unknown scene and object using methods such as nearest neighbour search.

LINEMOD[5] is an example of a template-based recognition method. It provides a framework for combining different modalities to create a template. In the original implementation, colour gradients and surface normals were combined to form templates. The templates are later used to recognize and localize objects in an unknown scene.

In [15], a global descriptor called Viewpoint oriented Colour-Shape Histogram is described. The shape descriptors are based on the relationship between points

and the centroid of the point cloud. Four features (two distances and two angles) for each point are measured and used to build the histogram.

In [7], the authors use colour descriptors, edge descriptors and shape descriptors as features for their fruit classifier. The shape descriptors include compactness, symmetry, local convexity and smoothness defined by Karpathy et al.[8], and image moment invariants defined by Hu [6].

Mustafa et al.[10] describe a multi-view object recognition system for a controlled industrial setting. They construct shape descriptors using 2D histograms of measures such as Euclidean distance, angles and normal distance between pairs of *texlets* (which describe local properties of a textured surface). Appearance descriptors are constructed using 2D histograms of the H and S components of texlet colour in the HSV colour space. Although they achieved a good recognition rate, small-sized objects were the cause of some of the miss-classifications.

The feature descriptors used in most of the methods tend to be bottom-up approaches. They try to capture a signature for objects using the distribution of features measured at the point level. In this paper, we describe global feature descriptors without using relations for individual points. Instead we try to capture the most salient features for an object by means of fitting bounding boxes, circles etc. Although some level of detail is still required in the point clouds, very small details are of less importance.

3 Approach

3.1 Segmentation

The service areas for RoboCup@Work tasks are flat surfaces on which objects are placed with a minimum distance of 2 cm between them [3]. The robot is positioned in front of the service area such that the arm-mounted 3D camera has a full or partial view of the workspace. A previously developed pipeline is used to detect the plane of the workspace, segment the points above the plane and cluster them based on Euclidean distance [2]. These point cloud clusters, which represent the objects on the workspace, form the input for the object recognition component developed here.

3.2 Data Collection

A set of point clouds for each object is collected for training and testing using the segmentation method explained above. Figure 2 shows some of the point clouds collected using an Asus Xtion PRO Live¹ RGB-D camera.

The objects are placed in various positions and orientations on the workspace while building the dataset. The camera is mounted on a stand at the approximate height and distance from the workspace as an arm-mounted camera on the robot. This allows the subsequently extracted features to be representative of all positions and orientations within the workspace. Hence, during runtime,

¹ https://www.asus.com/3D-Sensor/Xtion_PRO/

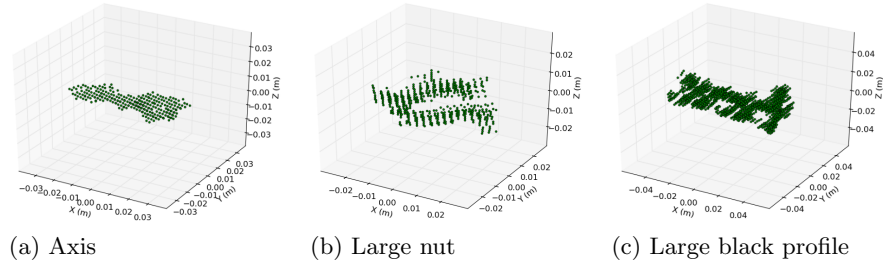


Fig. 2: Sample point clouds for a) Axis b) Large nut and c) Large black profile

the camera needs only to be approximately positioned in front of the workspace. The point clouds are translated to be centred at the origin and rotated such that the x, y and z axes align with the first three principal axes (retrieved using principal component analysis (PCA)) of the point cloud. This renders the extracted features invariant to the original pose of the object. Since the perceived colour of the objects is partially dependent on the lighting, it is expected that the point clouds are collected in the environment in which the objects will be used.

The set of point clouds used for training and testing are available online²

3.3 Features

Size and colour are the most salient features observable in point clouds. Additionally, circularity and the distribution of mass about the longitudinal axis also allow us to differentiate between a large set of various objects. Keeping this in mind, the following features are extracted from each object point cloud:

Bounding box The oriented bounding box of the points is calculated and returns the length, width and height of the point cloud and hence describes the size of the object.

Colour Since the colour of the objects are more or less homogeneous, only the mean and median colour are calculated. The red, green and blue channels of the colour component of each point are represented as a single floating point number as in PCL. The median and mean colour of the point cloud are calculated using this floating point representation.

Point cloud size The number of points in the object point cloud is indicative of the size of the object but is also dependent on the distance of the object from the camera. However, since the distance of the camera from the objects does not change drastically, this feature is also considered.

² https://github.com/sthoduka/atwork_object_dataset

Circularity Although the bounding box captures the size of the object, it treats every object as a rectangular cuboid. Since cylindrical objects such as nuts, bearings and bushings are common in industrial settings, the circularity of an object is an important feature as well.

- **Mean circle radius:** In order to measure the circularity of an object, a circle is fit on the x-y plane of the point cloud based on the mean squared distance of all points from the centre. The radius of this circle is indicative of the size of the object.
- **Radial density distribution:** Points are projected onto 36 equal segments of the circle to form a radial histogram. This distribution describes how round an object is. As seen in Figure 4, cylindrical objects (such as the nut) have a more uniform distribution whereas the distribution for longitudinal objects (such as the bolt) is more skewed along the principal axis. The radial density is calculated as

$$\frac{\sum_{j=1}^N \frac{k_j}{\max k}}{N} \quad (1)$$

where N is the number of bins in the histogram k .

A comparison of the radial density distribution for objects that are circular and non-circular in the X-Y plane is shown in Figure 3.

- **Outlier/inlier error ratio:** The outlier error to inlier error ratio is calculated as

$$\frac{\frac{\sum_{j=1}^{N_o} dist(po_j)}{N_o}}{\frac{\sum_{k=1}^{N_i} dist(pi_k)}{N_i}} \quad (2)$$

where po and pi are the points outside and inside the circle, N_o and N_i are the sizes of each set of points and $dist(x)$ is the distance of point x from the circumference of the circle. This ratio measures the hollowness of the object, with objects such as the nuts having a higher ratio compared to the motor.

Distribution of mass along principal axis Almost all of the longitudinal objects have an identical cross-section along their principal axis with the exception of the bolt and the axis. In order to differentiate these objects from the rest, the same circularity features, radius, radial density and outlier-inlier ratio, are calculated on eight slices along the principal axis. This adds an additional 24 features to the set.

Centre of mass offset Another feature considered is the offset between the centre of mass and the geometric centre of the object. This offset is higher for objects such as the bolt and axis which are not symmetric about the y-z plane. Figure 5 visualizes the bounding box, the circle fit on the x-y plane and the circles fit on the slices. The thickness of the visualized circles is proportional to the radial density. Although the small black profile and the bolt are very similar

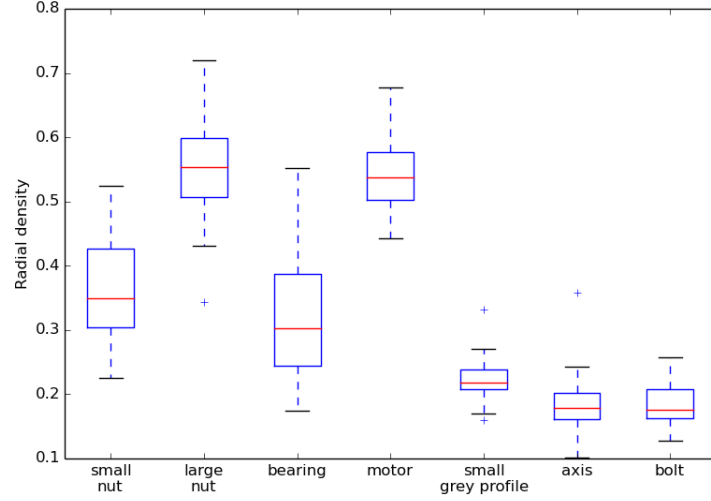


Fig. 3: Radial density distribution on the X-Y plane for cylindrical and non-cylindrical objects

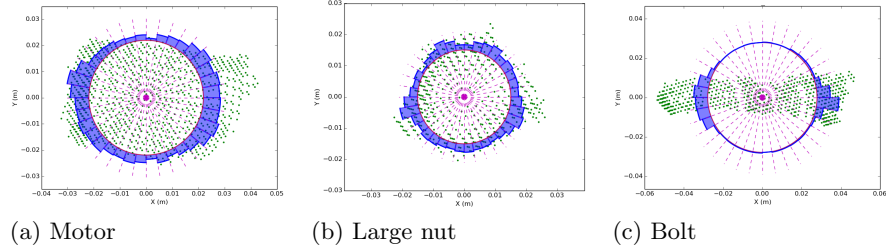


Fig. 4: Radial density distribution for a) Motor b) Large nut and c) Bolt

(similar bounding box, colour, mean circle radius etc.), the cap of the bolt is clearly identifiable by the larger circle compared to the similar-sized circles in the profile. Figure 6 shows the distribution of circle radii for the end slices and the remaining slices in the middle for the two objects. The larger range of radii for the bolt at the ends is likely to improve the classification between these two objects.

3.4 Training

A set of point clouds was collected for all objects in Figure 1 and was split into training and test data. A total of 34 features was extracted from the training data set. Various combinations of features, as described in Section 4, were considered in order to compare the impact of the different features on the classification rate. The feature set was standardized and used to train a multi-class support vector machine (SVM) classifier [11] with a radial basis function kernel.

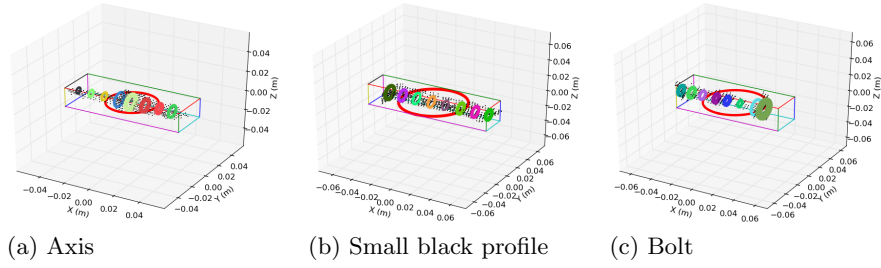


Fig. 5: Bounding box and mean circle features for a) Axis b) Small black profile and c) Bolt

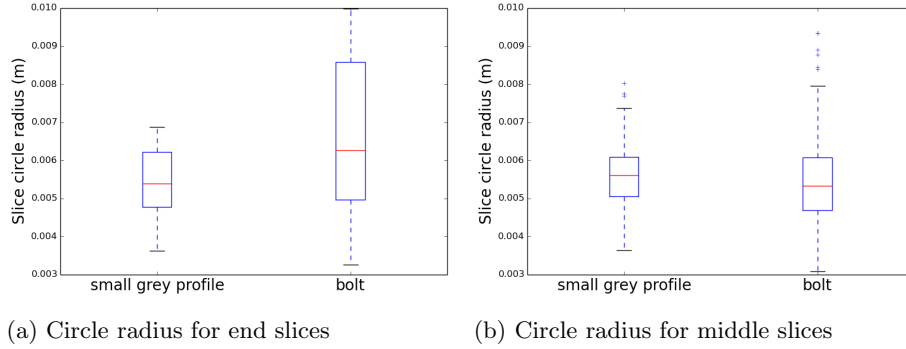


Fig. 6: Differences in slice circle radii for small black profile and bolt

3.5 Testing

In order to test the classifier, feature vectors are calculated on the test data and input to the classifier which returns a list of probability estimates for each object. The object with the highest probability is selected and a threshold is applied to increase the confidence of classification. If the probability is below the threshold, the object is said to be unclassified.

4 Results

In order to test the effectiveness of different features on the classification rate, the features were split into four categories described in Table 1. Six different combinations of feature categories were used to create different classifiers. The results of the different classifiers are presented in Table 2. A probability threshold of 0.5 was used to discard low-probability classifications (indicated as *unclassified* objects). Although using the probability threshold reduces the overall true positive rate, it lowers the false positive rate as well. However, the consequence of an incorrect classification is not as easy to fix as not recognizing an object at all. For example, if the robot does not recognize an object, it can attempt to

Table 1: Feature categories

Category	Description	Features
A	Bounding Box	length, width, height
B	Colour	median and mean colour
C	Point cloud size	number of points
D	Mean Circle	centre of mass offset, radius, radial density, outlier-inlier ratio for the x-y mean circle and circles along the principal axis

Table 2: Overall classification results using different combinations of features.

Features	True Positive Rate	False Positive Rate	Unclassified
A, B	0.88	0.04	0.08
A, B, C	0.9	0.03	0.07
A, B, C, D	0.89	0.03	0.08
A, B, D	0.89	0.03	0.08
B, C, D	0.86	0.04	0.1
B, D	0.86	0.04	0.1

A - Bounding box, B - Colour, C - Point cloud size, D - Mean circle

view the object from a different angle to try again. If the robot were to transport an incorrect object, it can cause a cascade of errors in subsequent tasks. The true positive rates for individual objects are presented in Table 3. In addition, the classification results using the local and global descriptor object recognition pipelines from PCL are presented for comparison. Signature of Histograms of Orientations (SHOT) [13, 14] with colour is used as the local descriptor and Ensemble of Shape Functions (ESF) [16] is used as the global descriptor. The poor performance of these methods is likely due to the small sizes of the clouds, making finding keypoints and the calculation of normals and surface properties harder. Since ESF does not consider colour, misclassifications between objects with only colour differences were considered correct.

Table 3: True positive rates for individual objects.

Object	A, B	A, B, C	A, B, C, D	A, B, D	B, C, D	B, D	SHOT	ESF
Small black profile	0.99	0.99	1.0	0.99	0.95	0.95	0.73	0.5
Small grey profile	1.0	1.0	0.96	0.95	0.81	0.8	0.65	0.41
Large black profile	1.0	1.0	1.0	1.0	1.0	1.0	0.42	0.33
Large grey profile	1.0	1.0	1.0	1.0	1.0	1.0	0.06	0.51
Bolt	1.0	0.97	0.98	0.98	0.87	0.85	0.68	0.68
Bushing	0.88	0.89	0.89	0.89	0.83	0.84	0.52	0.16
Small nut	0.31	0.32	0.48	0.5	0.42	0.43	0.47	0.0
Large nut	0.99	0.99	0.97	0.97	0.96	0.96	0.98	0.77
Bearing box	1.0	1.0	1.0	1.0	1.0	1.0	0.17	0.23
Bearing	0.74	0.73	0.65	0.65	0.58	0.61	0.34	0.02
Axis	0.66	0.8	0.75	0.77	0.76	0.78	0.1	0.1
Distance tube	0.78	0.91	0.77	0.75	0.74	0.74	0.0	0.8
Motor	0.8	0.89	0.93	0.92	0.93	0.91	0.75	0.24
Red container	0.99	1.0	1.0	1.0	1.0	1.0	0.25	0.3
Blue container	1.0	1.0	1.0	0.99	0.99	0.99	0.94	0.4

A - Bounding box, B - Colour, C - Point cloud size, D - Mean circle, SHOT - Signature of Histograms of Orientations, ESF - Ensemble of Shape Functions

The larger objects, such as the profiles, containers and bolts, are recognized with high accuracy. The small nut, bearing and distance tube have low classifications rates, likely due to their similarity. The misclassifications show that these are often confused with each other.

Introducing the mean circle features improves the recognition rate of the small nut, but marginally decreases the rate for the distance tube and bearing. The mean and median colour successfully classifies the identically shaped profiles and containers.

It is surprising that the point cloud size significantly increases the recognition rate of objects such as the axis and distance tube. It is, however, the least generalisable feature since it is dependent on the camera resolution and distance between the object and camera.

It is observed that adding more features is not always better. Adding irrelevant features increases the likelihood that the classifier over-fits to the training data. This makes the classifier less generalisable and it performs poorly on new data. A minimal set of features that are able to distinguish between objects should be selected.

5 Conclusions and Future Work

The proposed features and classifier are able to identify some of the objects with a high accuracy, but perform poorly for some of the smaller objects. The features, although designed based on the objects defined for RoboCup@Work, are sufficiently general that they can be applied to objects of the same class as those presented here. However, if variations of the some object classes (such as profiles) are present, an additional classification method may be required to distinguish between variants. It is trivial to add more features to the classifier if there is a need. However, care must be taken not to over-fit the classifier to the training data. The addition of 2D image features such as corners, edges and contours is a possible improvement to this method. With the continuous improvement of RGB-D cameras, the quality of the point clouds are expected to improve as well. Consequently, the performance of the method is also likely to improve.

Acknowledgements We gratefully acknowledge the continued support of the RoboCup team by the b-it Bonn-Aachen International Center for Information Technology and the Bonn-Rhein-Sieg University of Applied Sciences.

References

1. Factories of the Future. "http://www.effra.eu/attachments/article/129/Factories of the Future 2020 Roadmap.pdf" (2013), [Online; accessed 08-Mar-2016]
2. Ahmed, S., Jandt, T., Kulkarni, P., Lima, O., Mallick, A., Moriarty, A., Nair, D., Thoduka, S., Awaad, I., Dwiputra, R., Hegger, F., Hochgeschwender, N., Sanchez, J., Schneider, S., Kraetzschmar, G.K.: b-it-bots RoboCup@Work Team Description Paper. In: RoboCup. Leipzig, Germany (2016), https://mas-group.inf.h-brs.de/wp-content/uploads/2016/01/tdp_b-it-bots_atwork_2016.pdf
3. Carstensen, J., Hochgeschwender, N., Kraetzschmar, G., Nowak, W., Zug, S.: RoboCup@Work Rulebook Version 2016. <http://www.robocupatwork.org/download/rulebook-2016-01-15.pdf> (2016), [Online; accessed 08-Mar-2016]
4. Dwiputra, R., Berghofer, J., Ahmad, A., Awaad, I., Amigoni, F., Bischoff, R., Bonarini, A., Fontana, G., Hegger, F., Hochgeschwender, N., Iocchi, L., Kraetzschmar, G., Lima, P., Matteucci, M., Nardi, D., Schiaffonati, V., Schneider, S.: The RoCKIn@Work Challenge. In: ISR/Robotik 2014; 41st International Symposium on Robotics; Proceedings of. pp. 1–6 (2014)
5. Hinterstoisser, S., Holzer, S., Cagniat, C., Ilic, S., Konolige, K., Navab, N., Lepetit, V.: Multimodal Templates for Real-Time Detection of Texture-less Objects in Heavily Cluttered Scenes. In: Computer Vision (ICCV), 2011 IEEE International Conference on. pp. 858–865. IEEE (2011)
6. Hu, M.K.: Visual Pattern Recognition by Moment Invariants. *information Theory, IRE Transactions on* 8(2), 179–187 (1962)
7. Jiang, L., Koch, A., Scherer, S.A., Zell, A.: Multi-Class Fruit Classification Using RGB-D Data for Indoor Robots. In: Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on. pp. 587–592. IEEE (2013)

8. Karpathy, A., Miller, S., Fei-Fei, L.: Object Discovery in 3D Scenes via Shape Analysis. In: Robotics and Automation (ICRA), 2013 IEEE International Conference on. pp. 2088–2095. IEEE (2013)
9. Kraetzschmar, G.K., Hochgeschwender, N., Nowak, W., Hegger, F., Schneider, S., Dwiputra, R., Berghofer, J., Bischoff, R.: RoboCup@Work: Competing for the Factory of the Future. In: Bianchi, R.A.C., Akin, H.L., Ramamoorthy, S., Sugiura, K. (eds.) RoboCup 2014: Robot World Cup XVIII, Lecture Notes in Computer Science, vol. 8992, pp. 171–182. Springer International Publishing (2015)
10. Mustafa, W., Pugeault, N., Kruger, N.: Multi-View Object Recognition Using View-Point Invariant Shape Relations and Appearance Information. In: Robotics and Automation (ICRA), 2013 IEEE International Conference on. pp. 4230–4237. IEEE (2013)
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830 (2011)
12. Rusu, R.B., Cousins, S.: 3d is Here: Point Cloud Library (PCL). In: Robotics and Automation (ICRA), 2011 IEEE International Conference on. pp. 1–4. IEEE (2011)
13. Tombari, F., Salti, S., Di Stefano, L.: Unique Signatures of Histograms for Local Surface Description. In: Computer Vision–ECCV 2010, pp. 356–369. Springer (2010)
14. Tombari, F., Salti, S., Stefano, L.D.: A Combined Texture-Shape Descriptor for Enhanced 3D Feature Matching. In: Image Processing (ICIP), 2011 18th IEEE International Conference on. pp. 809–812. IEEE (2011)
15. Wang, W., Chen, L., Chen, D., Li, S., Kuhnlenz, K.: Fast Object Recognition and 6d Pose Estimation Using Viewpoint Oriented Color-Shape Histogram. In: Multimedia and Expo (ICME), 2013 IEEE International Conference on. pp. 1–6. IEEE (2013)
16. Wohlking, W., Vincze, M.: Ensemble of Shape Functions for 3D Object Classification. In: Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on. pp. 2987–2992. IEEE (2011)