Learning and Detecting Objects with a Mobile Robot to Assist Older Adults in Their Homes

Markus Vincze(✉), Markus Bajones, Markus Suchi, Daniel Wolf, Astrid Weiss, David Fischinger, and Paloma da la Puente

Technische Universität Wien, Vienna, Austria
vincze@tuwien.ac.at

Abstract. Older adults reported that a robot in their homes would be of great help if it could find objects that users regularly search for. We propose an interactive method to learn objects directly with the user and the robot and then use the RGB-D model to search for the object in the scene. The robot presents a turntable to the user for rotating the part in front of its camera and obtain a full 3D model. The user is asked to turn the object upside down and the two half-models are merged. The model is then used at predefined search locations for detecting the object on tables or other horizontal surfaces. Experiments in three environments, up to 14 objects and a total of 1080 scenes indicate that present detection methods need to be considerably improved to provide a good service to users. We analyse the results and contribute to the discussion on how to overcome limited image quality and resolution by exploiting the robotic system.

Keywords: Robot object learning · 3D object modelling · RGB-D model · Object detection in clutter

1 Introduction

Recently several vision methods have been used on mobile assistive or companion robots. These can be summarised to fall into three groups [1]: face, gesture and posture recognition methods for interacting with the user, navigation methods beyond using laser but rather RGB-D sensors to cope with the truly 3D environment, and methods to recognise objects.

It is interesting to note that most of the robots operate in care facilities. First robots have been operated either locally or remotely and did not possess autonomous navigation capabilities. Only recently a few robots moved out into the homes of users, e.g. [2,3]. The big step forward in recent work is that the robot should be at least partially autonomous in the user's home. So far robots have been operated remotely or only for very few tasks in a home during limited time of user trials, for example in [4–6]. It was pointed out by the researchers that the autonomous navigation capability would be of high importance.

Autonomous navigation in user homes enlarges the possible set of assistive functions. Experience has shown and user studies have confirmed that a modular

© Springer International Publishing Switzerland 2016
G. Hua and H. Jégou (Eds.): ECCV 2016 Workshops, Part II, LNCS 9914, pp. 316–330, 2016.
DOI: 10.1007/978-3-319-48881-3_22

approach with customisable features would most suitably satisfy the heterogeneous group of older people who could benefit from the use of a mobile robot in their homes. Example functions that could use computer vision methods are detecting emergencies, adaptive robot behaviour depending on user behaviour, picking up objects from the floor or other locations, or the detection of objects.

In workshops with older adults, users indicated that it would be a very useful function if the robot could find and detect objects. They reported to search relatively often for a few typical objects such as handbag or mug.

To realise this demand, we developed a procedure to detect objects that are of interest to users. This demand generalises to object search and delivery scenarios. To implement such an assistive function for robots needs three basic robot and vision capabilities.

1. Learn about the object of interest that the user wants to be detected and create a model for later usage.
2. Detect the object using the learned model.
3. Grasp the object and deliver it to the user either in the gripper or in a storage tray on the robot.

In this paper we report on the first two capabilities. Object grasping has been shown elsewhere already and object learning and detection are the core abilities for a robot in an object search and deliver scenario. For object learning and detection in a home setting there are two specific challenges that need to be tackled.

1. *Autonomous object learning:* The objects of interest will vary for every user. Hence, it is necessary to learn these objects. In a beginning phase an adviser or care person could assist, but assuming a wide use of service robots, this would not be feasible. Consequently, a method is needed that allows the user herself to teach the robot which objects are of her interest and need to be detected.
2. *Variety of home settings:* the detection procedure needs to be able to cope with detecting everyday objects that have very different types and it needs to find them under the largely varying conditions in a home environment. This conditions include but are not limited to the ambient illumination situation and that objects are typically not standing alone but rather found in cluttered scenes.

Further requirements such as detecting good search positions in the first places or an autonomous detection of such search location are beyond the scope of this work. We will focus on these two functions that are challenging in themselves.

The contribution of this paper is an approach that first lets the user model an object together with the robot and that then uses the learned model for detecting the object at specified search locations. To the best of our knowledge this is the first time that a user will trigger the learning procedure and conduct the procedure to acquire a full 3D model of the object of interest. Figure 1 shows the robot with the turntable that is used for object learning. The detection

procedure itself capitalises on a mix of well established methods to combine colour and depth information of the learned models for object detection. The method of acquiring the model of a specific object will be made available open source.

Fig. 1. The robot used for learning object models and then using the models for object detection. The robot is shown during the acquisition of the model of an object of interest. It uses an active robot head to direct an RGB-D camera towards the object. The object is placed on a turntable including natural script on its side walls (masked for reasons of anonymity). The user initiates the robot learning procedure and then the robot guides the user through the necessary modelling steps. The processing steps are executed autonomously.

The second contribution of the paper is to evaluate the learning and object recognition method in a robotic use case and scenario. We evaluated the method by presenting the robot with real-world scenes in three different environments. This includes that the robot autonomously navigated to the target locations. The intention is to learn the difficulties in real-world settings and to propose further work to render vision methods for assistive robots more and more robust.

The paper is structured as follows. After reviewing related work on model learning, we introduce the robot system approach to learning an object model and recognising the models in Sect. 2. Section 3 describes how the learned model is used in the robotic search procedure to detect the object. And Sect. 4 summarises the results of the experiments, analyses the problematic cases and discusses possibly improvements.

1.1 Related Work

Object modelling typically involves steps to accurately track the moving camera, segment the object from the background, and post-processing such as global camera pose optimisation and surface refinement. Approaches to learn models of objects either use distinctive features or the shape of the object in an iterative optimisation approach.

Regarding distinctive features, the most well-known methods are the Scale Invariant Feature Transform (SIFT) [7]. The feature points are used to find correspondences of object points in image pairs. This enables the registration of RGB-D images for modelling an object or scene in general. For example, in [8] the authors developed a Visual SLAM (Simultaneous Localisation and Mapping) approach that tracks the camera pose and registers point clouds in large environments. Distinctive points can also be used to directly reconstruct models for object recognition. For example, Collet et al. [9] register a set of images and compute a spare recognition model using a Structure from Motion approach.

Another type of method to acquire a model of an object from multiple images is based on the well established Iterative Closest Point (ICP) algorithm. For example, Huber et al. [10] as well as Fantoni et al. [11] focus on the registration of unordered sets of range images, while Weise et al. [12] track range images and propose an online loop closing approach.

Objects are typically learnt by using a turntable, e.g., [13]. There are also a few works that model objects in the hand of a robot. In [14] the authors propose a robotic object modelling approach where the object and the robotic arm are tracked with a variation of the articulated ICP approach. [15] tracks the target object including a loop closure for adjusting the model points after a full 360 degrees rotation. And the authors in [16] proposed an efficient SLAM-based registration method. Object models are built by selecting a volume of interest, defined by a user as an input mask in one image, plus the height above the support plane.

We extend these methods by allowing the user to handle and drive the modelling steps. With this we make sure that constraints due to the modelling methods are handled by the robot system rather than an expert user as in the works above. There is no need to select an object region or other expert input. It is also not necessary to control the distance of the object to the camera, to segment objects from the background, and tracking uses fiducial markers. Furthermore, we directly link the learned model to the object detection step. Regarding object detection, methods are numerous and a full review goes beyond the scope of this workshop paper. Existing object detection methods consider the case where a database of trained objects is used to match it with sensor data. Typically, the systems focus on individual algorithms that only work on objects with specific object characteristics, e.g., point features for 3D opaque objects [17], visual keypoint descriptor based systems like MOPED [9] for textured or [18] for translucent objects. Users neither nor not want to know about object features

or characteristics. Hence, we will use a method that combines known detection methods for object detection.

2 Learning and Detecting Objects of Users

In the spirit of an assistive robot, learning of objects must be interactive. We implemented this on our robot depicted in Fig. 1. To overcome practical issues of object learning as indicated above, we devised a turntable that is mounted within the robot body and that the robot can extract for this purpose. Earlier trials with asking the user to put the turntable into the hand of the robot failed since it is then difficult to obtain a repeatable location of the turntable in the robot hand. In the following we outline the learning procedure.

First, the user has to call the robot to bring it in a position close to the user. The user initiates the task of learning an object either by pressing the button on the touch screen or a verbal command. The robot moves slightly from the user using the depth image from the head to have sufficient clear space to be able to extend its arm. It then grasps the turntable located on its body and presents the turntable to the user in a position such that it can be reached conveniently.

In the next phase, the robot guides the user through the steps of learning the object model. First, the robot asks the user to place the object on the turntable. It then rotates executes a full rotation with the turntable while acquiring images from its head RGB-D sensor. The robot then asks the user to turn the object upside down and repeats the procedure. Finally the robot asks the use to take the object from the turntable and it restores the turn table into the robot body. This procedure takes about three minutes with the robot, where the arms moves particularly slowly when it is retrieved and restored to make sure the arm moves safely and will not hit the user or the environment.

As Fig. 1 shows, we designed a squared turntable which enables robust and accurate camera pose tracking relative to the turntable. Hence, any kind of object regardless of its texture or shape can be learned. Next, we summarise the model learning method using the acquired images.

2.1 Learning Object Models on a Turntable of the Robot

Given the positions of the turntable from tracking its pose, object learning is based on RTM - Toolbox for Recognition, Tracking and Modelling of Objects presented in [19] and available on-line (http://www.acin.tuwien.ac.at/?id=450). It operate as follows. First RGB-D images are captured and the camera pose is tracked with respect to the region of interest (ROI) covering the object and the squared turn table. Since the robot pose is known this ROI can be easily set using a depth segmentation around the known pose of the turntable.

Two algorithms, namely an image key point based pose tracking pipeline and an Iterative Closest Point (ICP) approach are implemented to estimate the camera motion. Both algorithms are state of the art and allow robust camera pose tracking. Additionally, we implemented the non-linear pose optimization

proposed by Fantoni et al. [11], which compensates the drift. A final filtering step using a weighted voxel grid inspired by KinectFusion [20] is used to subsample and smooth the reconstructed object point cloud. Results of learned RGB-D models are shown in Fig. 2. It shows rendering of eight objects from the textured 3D model.

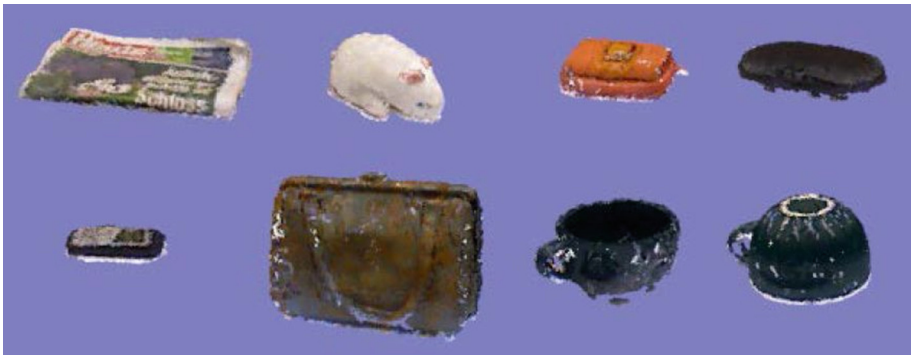

Fig. 2. Examples of models obtained from a set of images when rotating the object. The models are stored as textured 3D point clouds.

The RTM toolbox for object modelling and recognition [19] works best with object distances of about 80 to 160 cm using the standard Kinect. It can also be used with other RGBD sensors. Since resolution drops significantly at larger distances, models will then not be as accurate but modelling is possible.

2.2 Detecting Objects Using the Learned Object Model

For object detection we adopt a method that recently solved several databases for 3D object recognition [21]. We select this method since it combines in an optimisation framework the advantages of local and global methods of object recognition and proved to be effective by fully solving several challenges in databases for the first time. This serves the purpose of handling methods that can detect different types of objects with different characteristics such as with and without texture as outlined above.

The method is based on a combination of different recognition pipelines, each exploiting the data in a different manner and generating object hypotheses that are later fused in a Hypothesis Verification stage [22] that globally enforces geometric consistency between model hypotheses and the scene. Such a scheme boosts the overall recognition performance as it enhances the strength of the different recognition pipelines while diminishing the impact of their specific weaknesses. Specifically, the currently implemented pipelines take advantage of the multi-modality of the RGB-D data:

- A semi-global 3D descriptor representing an extension of CVFH approach [23] based on the colour, shape and object size cues. Regarding the segmentation stage required by the semi-global pipeline, we adopted the standard plane segmentation methods as available in the Point Cloud Library [24].
- A 2D local descriptor, SIFT [7], which is able to generate object hypotheses with associated 6 DOF (Degrees of Freedom) pose by back-projection of the 2D keypoint locations into the 3D space.
- A 3D local descriptor, SHOT [25] aimed at establishing correspondences between model and scene surface patches.

Given the object models as learned above, also detection works best in distances up to two metres. As will be explained in the procedure below, we used the proper distance to define appropriate search locations. The view angle is given due to the fixed height of the robot and a good angle down onto the table. Note, that object orientation can still be random since the full viewing sphere od the object has been model. An evaluation of object detection over distances is part of future work.

3 Procedure for Robotic Object Detection

Ideally the robot is able to search for the object of interest in any environment. However this would mean to autonomously create a sequence of view points that cover the full 3D environment. Such methods are not yet available [26]. Using the human model, the search should also take into account previously seen object locations and rooms, semantic information on where typically a specific object is located, or contextual information from the room structure that may bais the object search.

The actual robot implementation used a simple search procedure, based on the optimization of a cost function. As a prerequisite, several "search locations" per room have to be defined in the map during the initialization phase. These search locations are defined with the users and comprise tables and shelves where to detect the objects of interest. This may seems as a restriction at first, but be aware that algorithmic solutions are difficult [26] and would still need to capture the semantics of rooms and objects, both open research topics. Hence, in a first practical approach to evaluate real-world object detection, this seems a fair approach until more advanced methods become available. What comes close to include these semantics is the approach taken in [27], where the labels of items in the rooms [28] are taken to generate search locations on the fly.

If an object has to be searched for, the cost function is evaluated for every search location. The locations are then sorted according to their corresponding cost, which yields the optimized search procedure for the object. The cost function takes several aspects into account, such that a good trade-off between the probability of the object being found at a search location and the time it takes to get there can be found. While the different locations are searched by the robot one-by-one, the probabilities of the object being there are permanently updated depending if the object has been found or not.

Moreover, a penalty term is added to locations which are in the same room as the user, as we assume that the object is most likely located in a different room which should therefore be searched first. If a room cannot be reached because the path is blocked, the costs for all search locations in that room are increased such that these locations are considered last during the search procedure.

This comes close to using a semantic scene segmentation algorithm. The purpose of such an approach is to generate a segmentation of the scene, visible by the head Kinect, into semantically meaningful parts, like floor, wall, table, shelf, etc., e.g., [29]. Using the additional knowledge of the semantic segmentation result, the search procedure for objects could be further automated and optimized, without the need of specifically labelled search locations. Instead, the robot figures out itself, where possible object locations might be (objects are most likely located on tables, shelves...) and where it has to navigate to in order to be able to detect them. The knowledge could be exploited even further, to generate complete semantic maps of the environment, allowing the user to send the robot to automatically detected places like the table in the living room.

The proposed method uses this knowledge only insofar as it fits planes as part of the fitting procedure. The selection is given by the pre-set search locations. This is also necessary to obtain an evaluation of the detection methods and not mix detection with semantic labelling or view point selection.

The scenario we evaluate is the request of a user to detect a specific object. An example is given in Fig. 3. The evaluation procedure takes the image and detection is run versus all stored object models using the global hypothesis verification framework [21]. In a last step it uses the generated object detection hypothesis in a verification step, where the model is fit to the data. If this fit gives a confidence above 95 percent, the detection is reported as successful. Other detections with lower probabilities are not considered since most of the time found to be not correct. A thorough evaluation of this initial procedure is future work. In the following we describe the experiments conducted in more detail.

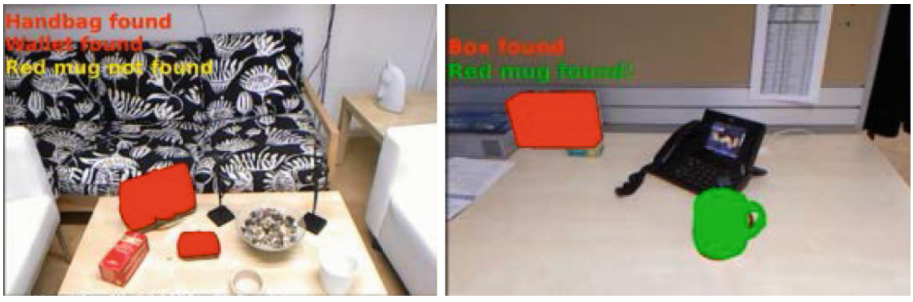

Fig. 3. Example of a search procedure. The task is to search for the red mug. When entering the room, two other locations are closer and searched first. Other known objects are found and their location is stored for future search operations.

4 Experiments

The tests for the object detection scenario were conducted in three environments that were built and furnished to resemble home settings. Systematic tests in a user home could not yet be made, since permitted time at the user sites was too short. The three test environments have been an office setting, where objects of daily use have been arranged in typically cluttered scenes (Env1), an ambient assisted living (AAL) laboratory that has been built with the purpose to resemble an older user home (Env2), and a living room setting specifically designed for testing assistive robot capabilities (Env3).

When conducting user trials with older adults, we defined search positions related to tables or boards where the users would typically place objects (anonymous, paper submitted). We replicated this procedure by defining four search positions in each of the three environments.

During the experiments for object detection the robot will navigate to the position that has the highest probability to find the object. This will use the knowledge of where the object has been seen before respectively knowledge about the typical rooms where a certain object is found. The robot will plan the shortest path to the selected search position to reduce unnecessary power usage.

The tests in user homes also indicated that objects are often found in clutter rather in the first test scenes such as Fig. 3. For our test we set up each scene with clutter. We also place one or multiple known objects on the table to check if more than one known object can be detected. We count objects only if they are within the robot's field of view at the search position. When the robot arrives at the search positions, it starts the recognition and records, which objects were recognized, which object were not recognized, and where an object was falsely reported (false positives). We conducted 10 trials for each search place. Figure 4 gives four examples scenes indicating the typical clutter, object occlusions, background illumination, and other effects that render object detection difficult in a natural environment.

We used 12 respectively 14 different objects in the three scenarios. For three environments this results in a total of 1080 trials where the robot autonomously navigated to one of the search locations and initiated the detection of all of the learned objects.

Table 1 summarises the results for the objects, the three environments and gives a summary. There have been no false positives–the method was tuned to not falsely report object detections. On the other hand, this leads to more objects going undetected. In the tables we report the number of successful detection out of the multiples of 10 trials each (different for each object between 10 and 50 trials). We report only positive detections, since the procedure as outlined above uses a verifications step that is very confident to select the correct object given the viewing of an object is satisfactory. An extension to work with less likely hypothesis is a good extension as indicated by the reviewers.

On first view the overall result of 52 % detection rate is not satisfactory. On the other hand, the task is indeed challenging and individual methods would rank much lower since specialised on one type and characteristics of objects

Fig. 4. Four example images showing the objects and the scenes with clutter. Note that objects can have very different viewpoints and scenes include cases with strong background that render detections difficult to infeasible.

only. A comparison to other methods is not directly feasible. As of today, there is no other modelling tool available that would allow a robot to obtain an object model and move in the environment to detect the object. A similar attempt has been made by learning objects from the robot in [30], but the robot navigates around objects rather than moves it with the robot arm and only a partial viewing sphere is captured.

Table 2 summarises results for the three different environments. Performance is similar. While the detection results of slightly more than half the objects are not impressive, this result was expected and reflects the present state-of-the-art in object detection methods when applied to the wild and in realistic settings. Furthermore, we know from the detection methods that feature-less objects such as toilet paper and water boiler are difficult to distinguish from the background. Similar difficulties give handbags if they are rather feature-less like the one used. On the other hand, larger objects with clear texture such as the ketchup, Mueller bottles, and tea box exhibit the expected satisfactory to good results.

Table 1. Summary of detection rate for each object and environment (Env# where # is 1 to 3; EnvAll refers to the sum of all three Environments). The numbers in the table give the successful detections of the target objects out of the 10 to 50 trials. In total 1080 object detections would have been possible.

Object	Envl		Env2		Env3		EnvAll		Rate
Asus Xtion box	0	10	20	20	12	40	32	70	0,46
Cisco telephone	20	40	21	40	30	40	71	120	0,59
Cleaning agent bottle	11	30	20	30	28	30	59	90	0,66
Felix ketchup bottle	10	10	17	20			27	30	0,9
Handbag	1	10	0	40	16	50	17	100	0,17
Muellermilch bottle banana	20	30	1	20	30	30	51	80	0,64
Muellermilch bottle choco	20	40	20	20	30	30	70	90	0,78
OpenCV book	25	30	13	40	1	10	39	80	0,49
Red mug with white dots	33	40	11	40	0	10	44	90	0,49
Strands mounting unit	10	20	1	20	0	10	11	50	0,22
Tea box	9	20	34	40	24	40	67	100	0,67
Toilet paper roll	0	30			0	30	0	60	0
Water boiler	0	10					0	10	0
Yellow toy car	29	40	20	30	21	40	70	110	0,64
Total	188	360	178	360	192	360	558	1080	0,52

Table 2. Summary of detection rate for the three environments.

Environment	# of different objects	# of trials	Detection rate
1	14	360	52,2
2	12	360	49,4
3	12	360	53,3
Total	14	1080	51,7

4.1 Discussion

When analysing the results, there are several factors that explain the many cases where objects are not detected.

– Limited dynamic range of the camera: often the robot enters a room and on the other side is a table. Looking against windows introduces highlights and reflections and renders objects dark. For robot navigation purposes we used high dynamic range cameras that improve but not resolve this case. Similarly, object detection methods need a mechanism to evaluate if the image in itself has feasible dynamic range and in principle allows to detect an object.

- Specific limits of sensitivity: sunlight through the window renders the depth image void. Similar to above, detecting these cases and reverting to methods that rely on other modalities will be the better robot system approach. Even better cameras and sensor will have specific characteristics that are better handled from a system perspective.
- Limited resolution of depth camera: non-textured objects are detected using the depth image. The resolution of the depth channel of present range cameras is much smaller than of colour images. Particularly with increasing distance, and this may well be the far end of the table, detection results deteriorate quickly. This cries for alternative sensors or using other approaches to overcome this issue.

For all the cases above, assistive vision approaches may profit most if the robot system exploits its mobility to select better view points and uses all its contextual knowledge about the environment for pruning hypothesis. The robot could exploit far and close range methods and purposively combine weak hypotheses by getting closer. It could detect other objects and use priors. And certainly, the robot will also profit from more advanced methods of reliably detecting objects in images, in particular objects of very different characteristics such as with and without texture, simple and complex shapes, or single and many colours.

Finally, the steady advance in camera technology brought us already to the level where we are right now. Hence, we can expect more and more advances and improvements in the near future from this side alone. Still, camera technology alone will not solve the case. The complementarity of methods and the exploitation of the robot system and the knowledge about the environment it has needs to be exploited in a much more rigorous fashion.

5 Conclusions

In this paper we investigated the scenario of assisting older adults with a method to learn their favourite objects and to detect the learned objects in a home environment. To this end we adapted a method to learn the object autonomously from the robot using a turntable and a clear procedure to guide the user through the learning method. We then spent considerable effort to run the robot autonomously to 1080 locations and view a given setting with small navigation uncertainties. Navigation in itself was fond to be accurate within a few centimetre and less than one degree. What we tested was the detection of up to 14 target objects using a set of four pre-set search locations with the small navigation uncertainties in three different environments.

We challenged the system by using target objects that had different characteristics with and without texture, single and multiple colours, and basic and more elaborate shapes. Adapting a method that globally optimises over multiple hypothesis we combines three detection methods to cope with these different object characteristics.

The results show that even a combination of methods achieves hardly satisfactory results. The analysis of the results indicates that camera properties

are not sufficient: both dynamic range and resolution are the main reasons for missed detections in a non optimal setting. Having control over object size, setting and image quality would render result much better. However these factors are difficult to control in an open home setting.

On the contrary, on of the intansion of this workshop paper is to contribute to the discussion on how vision methods that are typically trained on an image database can be made more suitable to the open settings on robots. The study indicates that present object detection methods are getting useful for assistive robots in a home setting but further work is needed.

Future work should have a look at the practical challenges posed in actual home settings. One such challenge is to locate objects that are visible but in the image the resolution, illumination situation or clutter do not allow present methods to detect the object. Following an idea presented already a decade ago in [31] we might use the cognitive power of humans to aid in detecting the target objects and learn from these detections. While in itself a cumbersome approach that will need a lot of user interaction, older adults indicated that they would be interested to help the robot. An ease-to-use interface with potentially indicated object hypotheses may be one option. We could here use less likely detections that may also include false positives, but users would be very quick to select the correct object. In this way the robot would learn both correct and false detections and could improve its object detection capability.

We see this only as starting a deeper discussion of the discrepancy between databse driven research and the open settings a robot would approach in homes. There is the need to make explicit the type of objects that can be handled by a certain method, discuss methods that integrate other methods, and—we think most of all—how to better exploit the contextual knowledge a robot has about a scene to improve detection results. Given a robot system it seems much more obvious to detect and exploit scene context rather than detecting it in an image alone.

Acknowledgements. The research leading to these results has received funding from the European Communitys Seventh Framework Programme FP7/2007-2013 under grant agreement No. 600623, STRANDS and No. 610532, SQUIRREL as well as No. 288146, HOBBIT.

References

1. Kragic, D., Vincze, M.: Vision for robotics. Foundations Trends Robot. **1**(1), 1–78 (2010)
2. Fazekas, G., Tth, A., Rumeau, P., Zsiga, K., Pilissy, T., Dupourque, V.: Cognitive-care robot for elderly assistance: preliminary results of tests with users in their homes. In: AAL Forum, Netherlands (2012)
3. Panek, P., Beck, C., Edelmayer, G., Mayer, P., Rauhala, M., Zagler, W.L.: Connecting AAL devices and systems to improve service delivery. In: AAL Forum: Broader, p. 2014. Romania, Bigger, Better - AAL solutions for Europe, Palace of Parliament, Bucharest (2014)

4. Bedaf, S., Gelderblom, G.J., de Witte, L., Syrdal, D., Lehnmann, H., Amirabdollahian, F., Dautenhahn, K.: Selecting services for a service robot - evaluating the problematic activities threatening the independence of elderly persons. In: ICORR (2013)
5. Ullberg, J., Loutfi, A., Pecora, F.: A customizable approach for monitoring activities of elderly users in their homes. In: Mazzeo, P.L., Spagnolo, P., Moeslund, T.B. (eds.) AMMDS 2014. LNCS, vol. 8703, pp. 13–25. Springer, Heidelberg (2014)
6. Glende, S., Conrad, I., Krezdorn, L., Klemcke, S., Krtzel, C.: Increasing acceptance of assistive robotics for older people through marketing strategies based on stakeholder needs. Int. J. Soc. Robot. **8**, 355–369 (2016)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
8. Endres, F., Hess, J., Sturm, J., Cremers, D., Burgard, W.: 3-D mapping with an RGB-D camera. IEEE Trans. Robot. **30**(1), 177–187 (2014)
9. Collet, A., Martinez, M., Srinivasa, S.S.: The moped framework: object recognition and pose estimation for manipulation. Int. J. Rob. Res. **30**(10), 1284–1306 (2011)
10. Huber, D.F., Hebert, M.: Fully automatic registration of multiple 3D data sets. Image Vis. Comput. **21**(7), 637–650 (2003)
11. Fantoni, S., Castellani, U., Fusiello, A.: Accurate and automatic alignment of range surfaces. In: Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), pp. 73–80 (2012)
12. Weise, T., Wismer, T., Leibe, B., Gool, L.V.: Online loop closure for real-time interactive 3D scanning. Comput. Vis. Image Underst. **115**(5), 635–648 (2011)
13. Dimashova, M., Lysenkov, I., Rabaud, V., Eruhimov, V.: Tabletop object scanning with an RGB-D sensor. In: SPME (2013)
14. Krainin, M., Curless, B., Fox, D.: Autonomous generation of complete 3D object models using next best view manipulation planning. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5031–5037, May 2011
15. Weise, T., Wismer, T., Leibe, B., Gool, L.V.: In-hand scanning with online loop closure. In: IEEE ICCV Workshop (2009)
16. Stueckler, J., Behnke, S.: Multi-resolution surfel maps for efficient dense 3D modeling and tracking. J. Vis. Commun. Image Represent. **25**(1), 137–147 (2014)
17. Aldoma, A., Marton, Z.C., Tombari, F., Wohlkinger, W., Potthast, C., Zeisl, B., Rusu, R.B., Gedikli, S.: Using the point cloud library for 3D object recognition and 6dof pose estimation. IEEE Robot. Autom. Mag. **9**, 80–91 (2012)
18. Lysenkov, I., Eruhimov, V., Bradski, G.: Recognition and pose estimation of rigid transparent objects with a kinect sensor. In: Proceedings of Robotics: Science and Systems, Sydney, Australia, July 2012
19. Prankl, J., Aldoma, A., Svejda, A., Vincze, M.: RGB-D object modelling for object recognition and tracking. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 96–103, September 2015
20. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, pp. 127–136. IEEE Computer Society (2011)
21. Aldoma, A., Tombari, F., Stefano, L.D., Vincze, M.: A global hypothesis verification framework for 3D object recognition in clutter. IEEE Trans. Pattern Anal. Mach. Intell. **PP**(99), 1 (2015)

22. Aldoma, A., Tombari, F., Di Stefano, L., Vincze, M.: A global hypotheses verification method for 3D object recognition. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 511–524. Springer, Heidelberg (2012)
23. Aldoma, A., Tombari, F., Rusu, R.B., Vincze, M.: OUR-CVFH – oriented, unique and repeatable clustered viewpoint feature histogram for object recognition and 6DOF pose estimation. In: Pinz, A., Pock, T., Bischof, H., Leberl, F. (eds.) DAGM and OAGM 2012. LNCS, vol. 7476, pp. 113–122. Springer, Heidelberg (2012)
24. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, pp. 1–4, 9–13 May 2011
25. Tombari, F., Salti, S., Stefano, L.: Unique signatures of histograms for local surface description. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6313, pp. 356–369. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15558-1_26
26. Andreopoulos, A., Hasler, S., Wersing, H., Janssen, H., Tsotsos, J., Korner, E.: Active 3D object localization using a humanoid robot. IEEE Trans. Rob. **27**(1), 47–64 (2011)
27. Bajones, M., Wolf, D., Prankl, J., Vincze, M.: Where to look first? Behaviour control for fetch-and-carry missions of service robots. In: Austrian Robotics Workshop(2014)
28. Wolf, D., Prankl, J., Vincze, M.: Enhancing semantic segmentation for robotics: the power of 3-D entangled forests. IEEE Robot. Autom. Lett. **1**(1), 49–56 (2016)
29. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 746–760. Springer, Heidelberg (2012)
30. Fäulhammer, T., Ambrus, R., Burbridge, C., Zillich, M., Folkesson, J., Hawes, N., Jensfelt, P., Vincze, M.: Autonomous learning of object models on a mobile robot. IEEE Robot. Autom. Lett. **2**(1), 26–33 (2016)
31. Makihara, Y., Takizawa, M., Shirai, Y., Miura, J., Shimada, N.: Object recognition supported by user interaction for service robots. In: 16th International Conference on Pattern Recognition, Proceedings, vol. 3, pp. 561–564 (2002)