Robust Manipulability-Centric Object Detection in Time-of-Flight Camera Point Clouds

S. Caraian and N. Kirchner
ARC Centre of Excellent for Autonomous Systems
University of Technology, Sydney
{sonja.caraian, nathan.kirchner}@uts.edu.au

Abstract

This paper presents a method for robustly identifying the manipulability of objects in a scene based on the capabilities of the manipulator. The method uses a directed histogram search of a time-of-flight camera generated 3D point cloud that exploits the logical connection between objects and the respective supporting surface to facilitate scene segmentation. Once segmented the points above the supporting surface are searched, again with a directed histogram, and potentially manipulatable objects identified. Finally, the manipulatable objects in the scene are identified as those from the potential objects set that are within the manipulators capabilities. It is shown empirically that the method robustly detects the supporting surface with ±15mm accuracy and successfully discriminates between graspable and non-graspable objects in cluttered and complex scenes.

1 Introduction

There is an increasing desire for robots to be both ubiquitous and useful in society, in areas ranging from manufacturing to service applications. In many cases, this requires robots to be capable of physically interacting with their environment, particularly in the form of object manipulation. Human environments are generally neither structured nor static: for example, tables and chairs are shifted, and objects are moved. Adaptable manipulation is only possible if the robot is equipped with a method of scene interpretation, enabling it to perceive the state of its environment [Bley *et al.*, 2006].

Accurate and robust perception is therefore vital for a robot to detect objects in the world which can be manipulated [Collet *et al.*, 2009]. There are two broad approaches to the challenge of object detection in such environments. In one method, the robot can have *a priori* knowledge of a set of objects, and an understanding of which it can manipulate. In the second, it can sense objects and have a means of identifying whether those objects can be manipulated. Since the world is full of a variety of objects that are continually changing, it is impractical to train the robot for the large number of items it may encounter. As a result, it is foreseeable that the robot may encounter objects that it does not recognise or cannot manipulate. It is therefore useful for the robot to be capable of observing an arbitrary, unknown object and determining whether it can physically interact with it.

It is reasonable to assume that many of the objects with which robots can physically interact are likely to be located on horizontal surfaces such as tables, desks and countertops. By exploiting these logical connections between elements of the world - locating these surfaces and searching on top of them - a robot can increase its chances of finding manipulatable objects. [Rusu *et al.*, 2009] propose a table detection approach which involves computing local surface normals of adjacent points within a time-of-flight (ToF) camera point cloud, then segmenting those points with normals approximately parallel to the world ground. This method is susceptible to the underlying sensor noise, which causes the apex of the mesh triangles to vary in the z-direction, inclining the normals. Small changes in coordinate points can result in large normal changes, affecting the method's reliability.

[Holz *et al.*, 2010] also employ local surface normals, but attempt to increase the method's surface detection reliability by adding the condition that each table point, in addition to having a normal parallel to the world ground, must be surrounded by a smooth surface to be categorised as lying on the table plane. However, material variation on the surface can offset data from the surface plane or, in the event of a highly reflective or translucent material, data may not exist at all. As neither knowing nor determining the material in advance are practical, these factors would affect the aforemen-

tioned surface detection techniques.

Regardless of the possibility of non-homogenous surface types, since there is a high probability that objects will be located on horizontal surfaces, [Rusu *et al.*, 2009] go on to find all objects above the detected surface, regardless of whether or not the robot is able to manipulate them. Similarly, [Holz *et al.*, 2010] assume the robot will be able to manipulate the detected objects. Methods of modelling "3D objects whose shape and location are unknown *a priori*" [Bone *et al.*, 2008], then grasp planning and execution were proposed by [Bone *et al.*, 2008] and [Richtsfeld and Vincze, 2008]. However, their systems are only tested on objects which are known to be manipulatable by the robots' grippers. [Saxena *et al.*, 2008] go a step further and focus on the task of identifying graspable locations on objects, for example the flute of a martini glass or handle of a mug. The assumption remains, however, that there will be a graspable point on the object.

To satisfy the need for graspable object detection, this paper presents a simple system which identifies manipulatable objects in ToF camera point clouds using a method of robust surface segmentation. A probabilistic method sensitive to surface characteristics is used to identify the most likely location of a horizontal surface, which is then searched for potential objects. Objects the robot is capable of physically interacting with are determined via defined dimensional object parameters based on the capabilities of the robot's manipulator.

The remainder of this paper is organised as follows. Section 2 discusses the fundamentals of ToF cameras in relation to complex point cloud error. The proposed graspable object detection and surface segmentation methods are detailed in Section 3, followed by the online object detection results in Section 4.

2 Time-of-Flight Camera Fundamentals

ToF cameras are a popular tool used to generate 3D models of the world for use in locating table tops and detecting objects due to their depth information, high frame rates, measurement range and accuracy and compact size. ToF cameras such as the SwissRanger 4000 (SwissRanger) by MESA Imaging, generate 3D distance data of a scene using the time-of-flight principle. LEDs in the SwissRanger emit amplitude-modulated infrared light to illuminate the scene. By measuring the phase shift between the emitted and reflected light, depth information is gained for all pixels in parallel. The phase shift is proportional to the object's distance. The sensor represents 3D data in a 2D image, where each of the 176 x 144 pixels represents a distance (as seen in Figure 1). ToF cameras' simplicity, distance algorithm efficiency and high speed make them ideal for real-time applications.

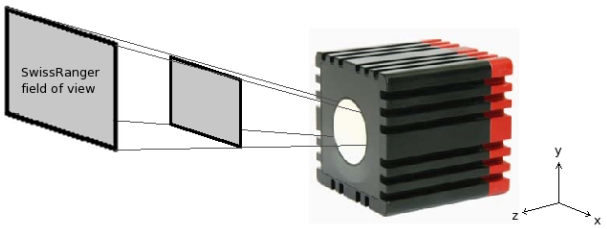

Figure 1: The SwissRanger field of view and default coordinate frame, which is used throughout this paper

It is well known that the measurement characteristics and accuracy of light-based sensors depend on the reflectivity of the objects in the surrounding scene [Kirchner, 2010; Kirchner *et al.*, 2009; May *et al.*, 2009]. The subject is also covered in the SwissRanger 4000 User Manual, which discusses two reflectivity issues which affect the measurement accuracy:

1. The ratio of back scattered light to incoming light
2. The angular distribution of the back scattered light (directed reflection versus diffuse reflection)

Diffusely reflecting materials (dull and matte surfaces such as paper) reflect the SwissRanger light with an intensity distribution that is independent of the observation angle. Directed reflecting materials (glossy surfaces such as glass and metal), however, can cause two issues:

1. The image can be saturated if the light is reflected directly into the sensor
2. If the reflected beam points away from the camera, the signal intensity can fall below a level suitable to deliver stable measurements

In addition to reflectivity, the 'back-folding' phenomenon (where points outside the sensor range are 'folded' back into the non-ambiguity range) and other environmental conditions can cause point cloud noise, including jump edges, where "phantom measurements occur at distance discontinuities, i.e. at the boundaries of surfaces partially occluding each other" [Holz *et al.*, 2010].

All ToF cameras are similarly affected by these issues. The adverse affects of this fundamental element of real-world sensing are significant, and as a result it is assumed that object and surface detection methods sensitive to noise will be affected by these sensor characteristics. Thus a robust method for interpreting the point cloud outputs to detect graspable objects and surfaces is necessary.

3 Manipulatable Object Detection

As previously mentioned, searching for manipulatable objects on flat horizontal surfaces is a logical approach,

as the method builds from *a priori* knowledge of an expected scene, rather than using a more complex solution.

The first step is to rotate the data in Euclidean space around the sensor's x-axis in order to align it with with the world coordinates. The transformation matrix seen in Equation 1 is utilised.

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad (1)$$

where the rotation angle θ is the tilt angle of the head.

With the data transformed from the sensor coordinate frame to the global coordinate frame, identifying surfaces is the next step.

3.1 Horizontal Surface Detection

Surface Plane Detection

The table detection technique takes advantage of the typical high point density on flat horizontal surfaces in gathered point clouds. Due to the large flat areas of horizontal surfaces, the associated points will cluster at the same height. The table top seen in the point cloud in Figure 2 (a), for example, accounts for over 45% of the total number of cloud points. It is therefore assumed that the maximum likelihood of a table will occur in the region of highest point density in the point cloud – in a typical scene, a table area will be much larger than other objects, such as lamps or chairs.

To detect the high density region, the search strategy was designed to be sensitive to the desired characteristics: a high point density parallel to the world ground. A histogram (h), as seen in Equation 2, with k bins is constructed along the cloud y-axis, such that the bins are parallel to any horizontal surfaces.

$$h = \sum_{i=1}^{k} n_i \quad (2)$$

The high density surface regions will have an apparent thickness of approximately 50mm, which is proportional to the ToF sensor noise. Assuming the noise has a Gaussian distribution, the most likely location of the surface is at the centre of the apparent thickness. A bin width (w) of 10mm, notably smaller than the apparent surface thickness, guarantees the table will be split into multiple bins.

The resultant histogram representation, shown in Figure 2 (b), gives a non-dimensional probability distribution illustrating the likelihood of a surface. The histogram bin with the highest point count is most likely to be the centre of the apparent thickness and real table height.

It is potentially the case that the field of view of the ToF sensor would result in it 'seeing' more of the floor than of a table. This could result in the system incorrectly detecting the floor as the horizontal surface most likely to be supporting objects. However, this is unlikely due to the characteristics of the search, which, in addition to identifying point clusters which are aligned with the sensor z-axis, is sensitive to high point density in such clusters. As shown in Figure 2 (a), the floor in the resultant point cloud has a significantly higher level of noise and error in the y-direction than the table surface. The probability distribution in Figure 2 (b) illustrates the flatter, wider peak at the location of the floor compared to the sharp table probability peak. This makes it unlikely that the floor will be incorrectly identified as the table.

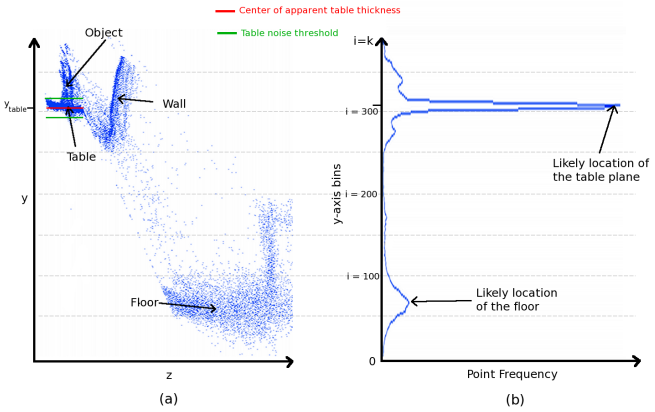

Figure 2: Point cloud viewed from the y-z plane and corresponding probability distribution along the y-axis, showing the likely location of the horizontal surface. The high floor noise is due to the relationship between reading distance and noise.

Determination of the table height is inherent to the search method: the probable table height is likely to be represented by the highest histogram peak. A table noise threshold (t) adds thickness to the plane, minimising the effects of noise due to surface irregularities and data rotation inaccuracies on the success of the detection. The table layer becomes the area between $y \pm t$, as shown in Figures 2 and 3.

The table height and maximum error are both functions of the bin resolution. Maximum error occurs if a bin boundary is located at the centre of the table's apparent thickness, and is equal to half a bin width. A small bin size therefore has the affect of reducing error in the calculated table height.

Surface Limit Detection

With the height determined, a representation of the surface edges is required in order to limit the search space during the object detection stage. The field of view of the ToF camera makes it likely that the point cloud surface

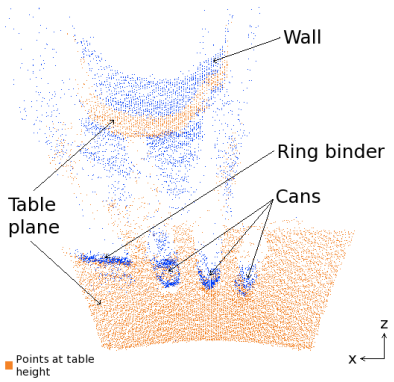

Figure 3: A typical scene as observed by the SwissRanger of three cans and a ring binder on a table near a wall. All points at the table height are highlighted in orange.

will be trapezoidal in shape, with the near and far edges aligned with the sensor x-axis. As shown in Figure 4, a rectangular bounding box aligned with the sensor x-z axes is used as an estimate of the surface boundaries.

It can been seen from Figure 4 that the rectangular box is an over-approximation of the real surface limits. However, in this case the object detection method searches only for objects supported by the horizontal surface. No objects can exist in the empty area between the bounding box and real surface, as they would be unsupported in physical space. The box is therefore a reasonable approximation of the surface.

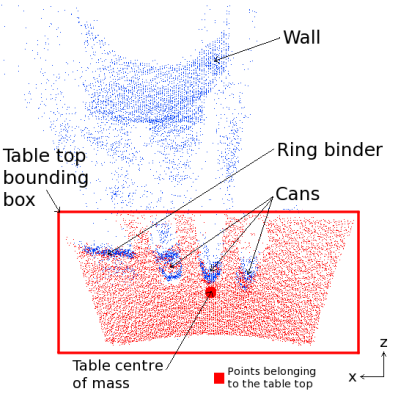

Figure 4: Point cloud showing the bounding box of the table top, with all points belonging to the table top highlighted in red

To calculate the parameters of the rectangular bounding box, the x, y and z components of all the points in the table plane are averaged together to find the centre of mass.

The heuristics of a number of scenes, both with and without background walls, showed a high percentage of the table plane mass was generally contained within the surface itself: a 90% mass threshold was found to be a reasonable approximation of the general proportion of table plane mass contained in the surface. The calculated centre of mass is therefore an acceptable estimate of the centre of mass of the observed portion of the table.

Once the centre of mass of the surface has been determined, the limits are required. The limits of the bounding box are grown until the proportion of plane mass contained within the box exceeds the mass threshold (as seen in Figure 4).

3.2 Object Detection and Manipulability Discrimination

At this stage the surface height has been determined and a representation of the surface limits has been developed. The next step is to detect objects which are supported by the horizontal surface.

The system disregards points outside the table plane and table bounding box and focuses on the table top, the area with the highest probability of containing a graspable object. In the first step of detecting multiple objects, a 3D, bivariate histogram is generated from the remaining points in the x-z plane using Equation 3 (as seen in Figure 5).

$$h = \sum_{i=1}^{k}(\sum_{j=1}^{l} n_{ij}) \tag{3}$$

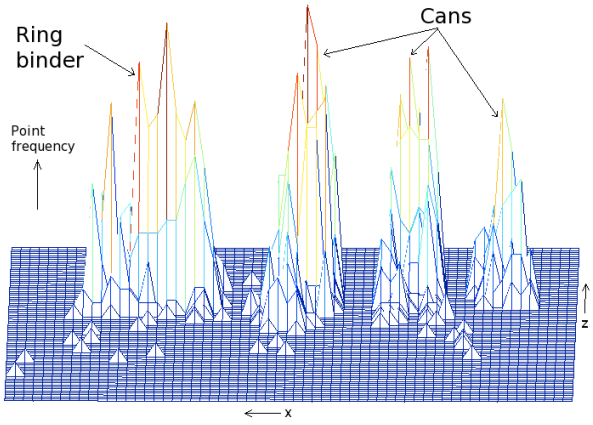

Figure 5: 3D, bivariate histogram showing the locations of the objects on the table top, including a sideways ring binder and three cans

The histogram is 'flattened' into an image where the pixel intensity is proportional to the bin count, showing the 'footprints' of any potential objects on the table. After converting the histogram from greyscale to binary using Otsu's method for adaptive grey level thresholding, MATLAB's 'regionprops' algorithm locates and determines the number of 'footprints', as well as calculating

the parameters of their bounding boxes (shown in Figure 6).

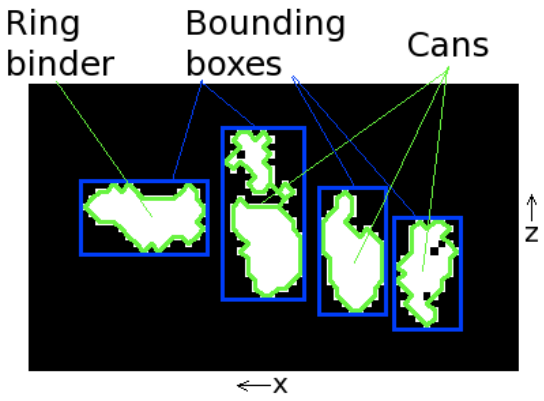

Figure 6: 'Flattened' histogram showing the object 'footprints' (outlined in green). Blob detection then identifies the bounding boxes (shown in blue).

Using defined dimensional parameters shown in Figure 7 – object width and height – which are based on the capabilities of the manipulator, the robot determines whether or not it can physically interact with the potential objects. Manipulatable width is limited by the maximum spread of the robot's gripper, and the height parameter excludes potential objects which, though they may fall within the width limits, are not manipulatable (such as poles or walls). Currently a depth parameter is not included in the dimensional check as depth cannot be reliably measured from a single observation (as shown in Figure 7 (b)). Depth criteria will be added to the method in future work.

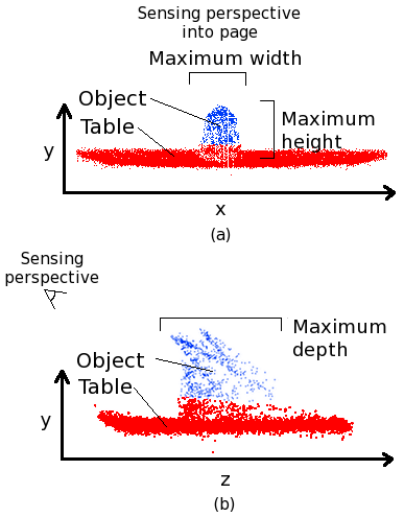

Figure 7: Dimensional parameters shown on a typical object point cloud

To determine whether each potential object satisfies the dimensional criteria, the object's corresponding bounding box is extended upwards above the table surface to form a search area surrounding the potential object point cluster in the 3D point cloud data. Carrying out the object cluster dimensional analysis in the original data rather than the 'footprint' histogram negates any inaccuracies introduced by the conversion of the histogram to binary. Using the point cluster centre of mass as the reference point, each object is checked against the dimensional parameters. To be classified as manipulatable, the object must satisfy both the dimensional conditions (Figure 8).

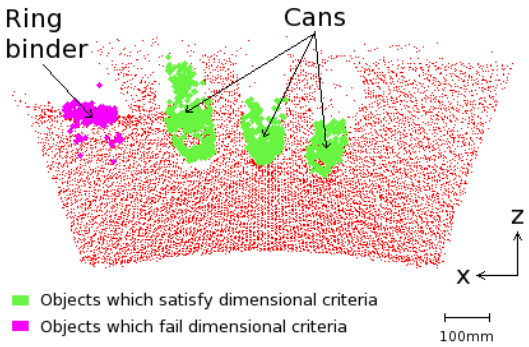

Figure 8: The green objects (cans) pass the dimensional tests, however the purple item (sideways ring binder) exceeds the maximum width

3.3 Summary

This section presented a robust manipulatable object detection system which uses a probabilistic search strategy sensitive to surface features to determine likely surface locations in the point cloud. A bounding box based around the approximate centre of mass of the visible portion of the table is constructed, and used to limit the object search space. To locate objects, a 3D, bivariate frequency histogram of the points above the table top is 'flattened' into an image where pixel intensity is proportional to bin count. A blob detection algorithm calculates the bounding boxes of each potential object 'footprint'. The final step determines which of the objects are manipulatable using dimensional parameters based on the robot gripper's capabilities.

4 Online Object Detection Results

This section presents indicative results of the system's performance. A number of online object detection tests were designed to verify the robustness and accuracy of the proposed manipulatable object detection and horizontal surface segmentation methods. In order to evaluate the effectiveness of the system, these experiments

were conducted with a variety of different objects in several arrangements. Three specific tests were run:

1. Non-homogenous horizontal surface detection
2. Discrimination of simple manipulatable objects on a homogenous surface
3. Discrimination of manipulatable objects in a complex scene

In all cases the manipulatable constraints were set to: height - *300mm* and width - *90mm*.

4.1 Robust surface detection

The first test was designed to test the system's robustness to detecting a horizontal surface with a variety of common materials on it. With the SwissRanger mounted on the RobotAssist robotic platform, a scan was taken of a desk covered by a number of relatively flat diffusely reflecting materials (leather wallet, fabric, white paper and desk surface) and directed reflecting materials (glass, shiny plastic in-tray, laptop computer, glossy magazine and metal pen), as seen in Figure 9.

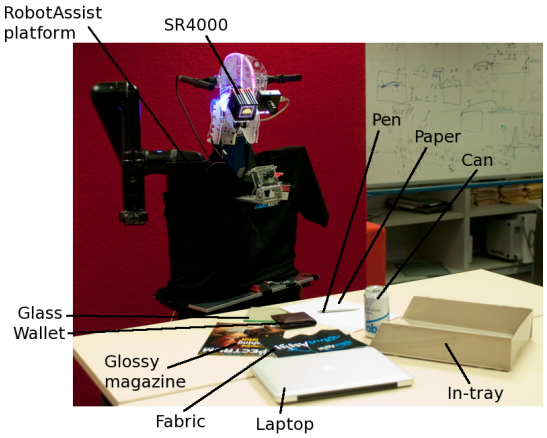

Figure 9: A typical desk with a variety of diffusely and directed reflecting surface materials

Figures 10 and 11 show the resultant point cloud and surface probability histogram. As can be seen in Figure 10 (a) and Figure 11 (b), the region of the laptop has a significant offset in the negative y direction, which is most likely due to the highly reflective surface of the polished metal material. The sharp change in y-coordinate values of the points on the boundary between the table and laptop would significantly effect horizontal surface detection methods which rely on the gradients of adjacent points being parallel with the z-axis. However, due to the nature of the search employed by our method, the large error does not significantly affect the detection of the horizontal surface plane (shown in Figure 10 (b)).

The horizontal surface plane was detected at a height of 735mm. This was compared with the true table height

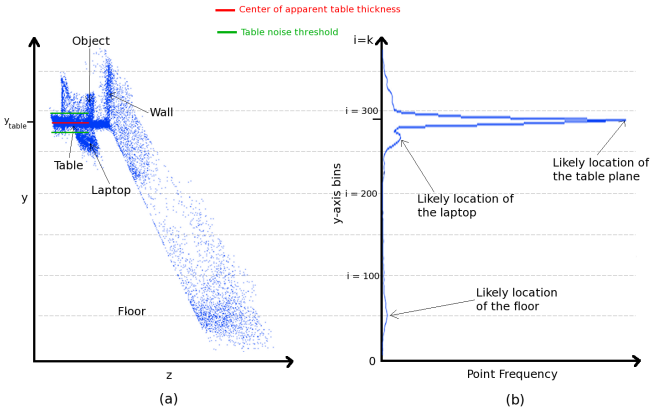

Figure 10: Test 1 point cloud viewed from the y-z plane and corresponding histogram along the y-axis, showing the likely location of the horizontal surface.

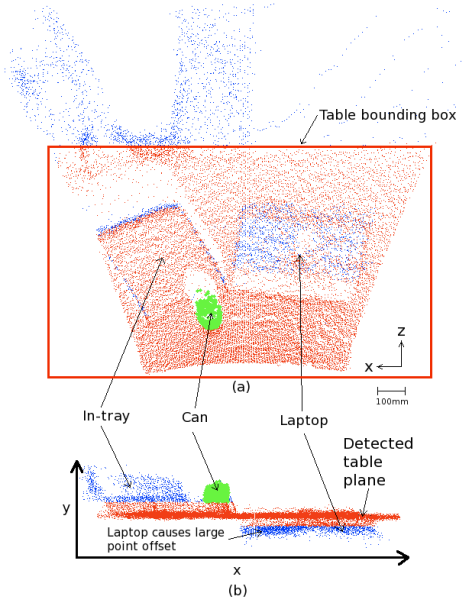

Figure 11: Point clouds showing the detection of a graspable can (highlighted in green) and the table plane on a horizontal surface (highlighted in red) with a large error due to the laptop

of 720mm, giving an accuracy of ±15mm, which correlates to the sensor distance accuracy and shows the robustness of the table height detection. The error could be affected by a number of factors including the sensor noise, the histogram bin width, the offset of the sensor from the ground, and the discrepancy between the reported and actual angle of the sensor to the ground.

Once the table plane was successfully identified, the system automatically calculated the parameters of the surface bounding box, shown in Figure 11 (a), robustly

cropping out the points not contained within the table surface and limiting the object search space to the table top.

The graspable can (highlighted in green in Figure 11) was then successfully located in the search space.

4.2 Identification of simple manipulatable objects on a homogenous surface

The second test was designed to demonstrate the detection and identification of arbitrary manipulatable objects in a cluttered scene. A number of relatively simple objects were placed in the SwissRanger's field of view on a homogenous surface, including a roll of sticky tape, a tea cup, a can, a clear water bottle, a bottle of Sprite, and a tin of coffee (illustrated in Figure 12).

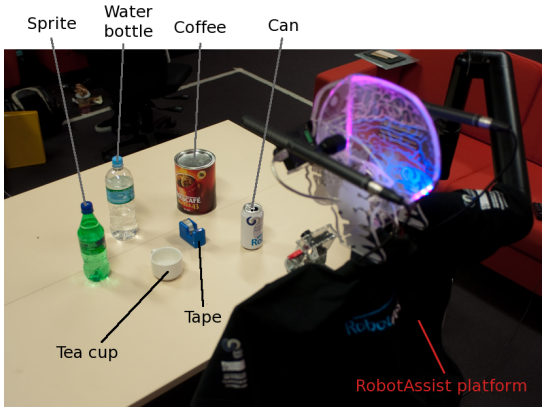

Figure 12: A desk with multiple objects, some of which can be manipulated by the robot

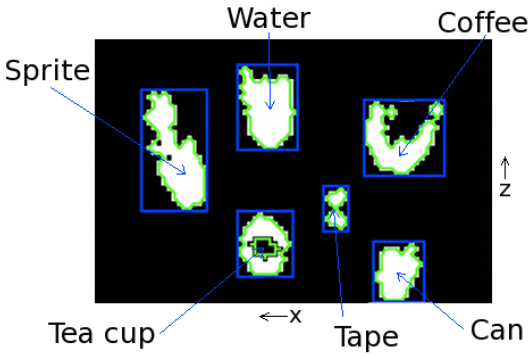

Figure 13: 'Flattened' histogram showing the object 'footprints' (outlined in green). Blob detection then identifies the bounding boxes (shown in blue).

Figure 13 shows the footprints of all items considered potential objects by the system. It can be seen that the table area was again successfully cropped, all of the items on the table were identified as potential objects and bounding boxes constructed around them. The coffee tin footprint illustrates that the spurious measurements in the jump edges, which are shown in the point cloud in Figure 14, are correctly filtered out when the histogram is converted from grayscale to binary. The histogram instead highlights the dense collection of points that will be found on the front surface of objects, from the sensing perspective, and indicates the real object locations.

Table 1 compares the real object dimensions with the detected ones. It can be seen that the average width error is ±14mm and height error ±35mm. Although the height error is sizeable, this parameter is less critical than width and is intended to prevent items such as walls and poles from passing the dimensional test. Furthermore, the larger error will, in most cases, not affect the successful identification of graspable objects. The system's robustness to noise is evident in the successful discrimination of manipulatable objects: the coffee (width *130mm*) and tea cup (height *55mm*) were found to fail the manipulability constraints, while the Sprite, water bottle, sticky tape and can were found as objects the robot could physically interact with, as shown in Figure 14.

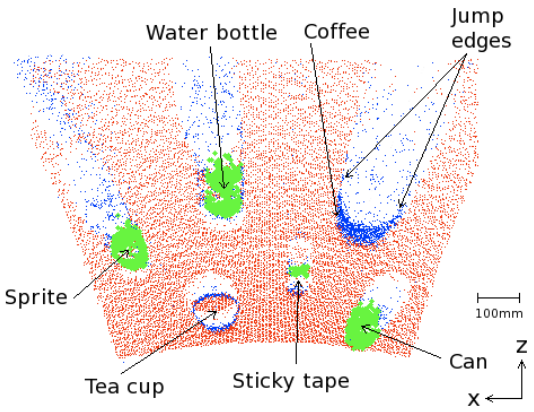

Figure 14: Point cloud with the manipulatable objects highlighted in green

4.3 Robust table detection and graspable object identification in a complex scene

The third and final test was an extension of the second test, designed to demonstrate the robustness of the object and table detection in a more complicated scene. Figure 15 shows how a number of items with large flat surfaces were placed on a typical desk, including a stack of books, a tissue box and a laptop, which has already been seen to create significant errors in the surface point cloud. Several manipulatable items (the ring binder –

Item	Actual Width	Measured Width	Actual Height	Measured Height
Can	60	76	130	110
Water	90	80	260	210
Sprite	70	85	235	200
Coffee	130	130	165	200
Tea Cup	90	85	50	55
Tape	40	50	65	70
Table	-	-	720	731

Table 1: Dimensional results from Test 2, showing the real and measured object dimensions. All dimensions are in mm.

now with its spine facing the sensor – water bottle and can) were also distributed on the desk.

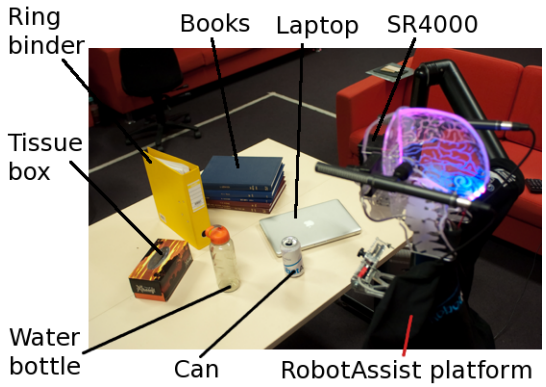

Figure 15: A cluttered desk with different surface materials, and graspable and un-graspable objects

Despite a high proportion of the desk surface being obscured from the sensor by the objects, the resultant point cloud in Figure 16 illustrates the successful location and cropping of the table (highlighted in red). Even with the erroneous readings caused by the cluttered objects, Table 2 illustrates that the table top, due to its high point density, still accounted for almost 63% of the total number of cloud points.

The successful surface identification enabled the system to identify the water bottle, ring binder and can (highlighted in green) in the search space as objects able to be manipulated by the robot. In the case of the ring binder, the system identified that the object was now in an orientation in which it could be manipulated.

4.4 Summary

These tests were designed to generate empirical results of the robustness of the object and horizontal surface detection methods. The first test illustrated the detection of a horizontal surface covered in a number of relatively

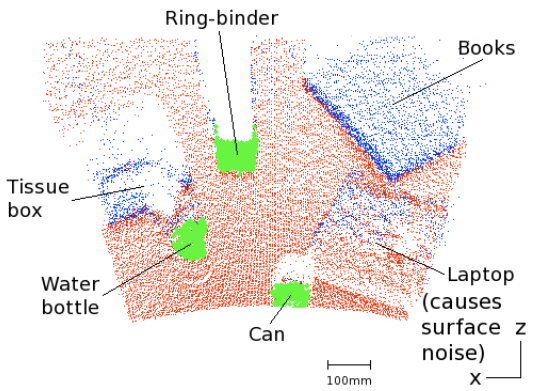

Figure 16: Point cloud of a cluttered desk surface, showing table surface and graspable objects

Total cloud points	25,344
Table plane points	16,365
Table surface points	15,889

Table 2: Point counts for Test 3

flat diffusely and directed reflecting materials, including accurate determination of the table height and boundaries in the presence of significant table noise and error. In the second test, the system successfully analysed a cluttered scene to detect simple arbitrary manipulatable objects, illustrating the method's robustness to object noise such as jump edges, and showing the accurate determination of object dimensions. In an extension of the second test, the final test presented a more complex scene including objects with large flat surfaces offset from the table height. Even with a high proportion of the desk surface obscured, the surface and object detection were both successful. From the experiments, it can been seen that the proposed methods are robust to a variety of environmental conditions.

5 Conclusions and Future Work

This paper presents a method for robustly identifying the manipulability of objects in a scene based on the capabilities of the manipulator. The method uses a directed histogram search of a time-of-flight camera generated 3D point cloud that exploits the logical connection between objects and the respective supporting surface to facilitate scene segmentation. Once segmented the points above the supporting surface are searched, again with a directed histogram, and potentially manipulatable objects identified. Finally, the manipulatable objects in the scene are identified as those from the potential objects set that are within the manipulator's capabilities.

The results presented demonstrate the suitability of the approach for the stated objectives. Experiments

have clearly demonstrated the system's ability to detect surfaces to an accuracy of ±15mm in the presence of significant reading errors due to material variations and sensor noise. Tests have also shown the correct identification and discrimination of manipulatable objects in cluttered and complex scenes using dimensional object parameters, including object width, which can be detected to an accuracy of ±15mm.

Future work will focus on extending the system to classify objects' material types, with a particular focus on estimating the weight of such objects based on their physical dimensions and material type, in order to determine if they can be manipulated by the robot. A depth parameter will also be added to the object dimensional check, and additional information will be determined, such as appropriate gripper approach angle. The system will be extended over multiple ToF camera scans in order to handle objects which become or are no longer occluded.

Acknowledgments

This work is supported by the ARC Centre of Excellence for Autonomous Systems, RobotAssist and the University of Technology, Sydney.

References

[Bley *et al.*, 2006] F. Bley, V. Schmirgel, and K.-F. Kraiss. "Mobile Manipulation Based on Generic Object Knowledge". In *The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 411–416, 2006.

[Bone *et al.*, 2008] G.M. Bone, A. Lambert, and M. Edwards. "Automated Modeling and Robotic Grasping of Unknown Three-Dimensional Objects". In *IEEE International Conference on Robotics and Automation*, pages 292–298, 2008.

[Collet *et al.*, 2009] A. Collet, D. Berenson, S. S. Siddhartha, and D. Ferguson. "Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation". In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 48–55, 2009.

[Holz *et al.*, 2010] D. Holz, R. Schnabel, D. Droeschel, J. Stückler, and S. Behnke. "Towards Semantic Scene Analysis with Time-of-Flight Cameras". In *In Proceedings of 14th International RoboCup Symposium*, 2010.

[Kirchner *et al.*, 2009] N. Kirchner, D. K. Liu, and G. Dissanayake. "Surface Type Classification With a Laser Range Finder". In *IEEE Sensors Journal*, volume 9, pages 1160–1168, 2009.

[Kirchner, 2010] N. Kirchner. "Exploiting Laser and Capacitive Ranging Sensors' Behaviour to Identify Material Type". In *PhD. Thesis*. University of Technology, Sydney, 2010.

[May *et al.*, 2009] S. May, D. Droeschel, D. Holz, S. Fuchs, E. Malis, A. Nüchter, and J. Hertzberg. "Three-Dimensional Mapping with Time-of-Flight Cameras". In *Journal of Field Robotics*, volume 26, pages 934–965, 2009.

[Richtsfeld and Vincze, 2008] M. Richtsfeld and M. Vincze. "Grasping of Unknown Objects from a Table Top". In *Workshop on Vision in Action: Efficient strategies for cognitive agents in complex environments*, 2008.

[Rusu *et al.*, 2009] R.B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz. "Close-range Scene Segmentation and Reconstruction of 3D Point Cloud Maps for Mobile Manipulation in Domestic Environments". In *IROS'09: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1–6, 2009.

[Saxena *et al.*, 2008] A. Saxena, J. Driemeyer, and A. Y. Ng. "Robotic Grasping of Novel Objects using Vision". In *The International Journal of Robotics Research*, volume 27, pages 157–173, 2008.