Pose Estimation of Texture-less cylindrical Objects in Bin Picking using Sensor Fusion

Mayank Roy[1], R. A. Boby[2], Shraddha Chaudhary[3], Santanu Chaudhury[4], S. Dutta Roy[5], S. K. Saha[6]

***Abstract*— We propose an approach for emptying of bin using a combination of Image and Range sensor. Offering a complete solution: calibration, segmentation and pose estimation, along with approachability analysis for the estimated pose. The work is novel in the sense that the objects to be picked are featureless and uniformly black in colour, hence existing approaches are not directly applicable. A key point involves optimal utilization of range data acquired from the laser scanner for 3-D segmentation using localized geometric information. This information guides segmentation of the image for better object pose estimation, used for pick-and-drop. We analytically assure the approachability of the object to avoid collision of the manipulator with the bin. Disturbance of objects caused during pick up has been modelled, which allows pickup of multiple pellets based on information from a single range scan. This eliminates the necessity of repeated scanning and data conditioning. The proposed method offers high object detection rate and pose estimation accuracy. The innovative techniques aimed at reducing the average pickup time makes it suitable for robust industrial operation.**

I. INTRODUCTION

A common bin-picking task involves object identification, pose and orientation estimation for a suitable pick-up, and subsequent emptying of the bin. Complications arise when a very large number of symmetrical and identical feature-less and texture-less objects (such as black cylindrical pellets) are randomly piled up in the bin, with arbitrary orientations and heavy occlusion. Fig. 1 shows an example of such a situation. The task at hand is to use a set of sensors with complementary properties (a camera and a range sensor) for pose estimation in heavy occlusion, accordingly orienting the manipulator gripper to pick up a suitable pellets, one-by-one. The manipulator is to avoid collision with the bin walls, and identify and characterise cases when the object is present in a blind spot for the manipulator. The ultimate aim is clearing of the bin, while minimizing the overall time required.

*This work was supported by grant towards setting up Programme for Autonomous Robotics Laboratory at IIT Delhi from BRNS, India.

[1] Mayank Roy is Research Fellow in Elec. Eng. Dept., IIT Delhi, India. mayank.ry@gmail.com

[2] R. A. Boby is Ph. D. Scholar in Mech. Eng. Dept., IIT Delhi, India. ribyab@gmail.com

[3] Shraddha Chaudhary is Ph. D. Scholar in Elec. Eng. Dept., IIT Delhi, India. chaudhary.shraddha18@gmail.com

[4]Santanu Chaudhury is Professor in Electrical Eng. Dept., IIT Delhi, India. santanuc@ee.iitd.ac.in

[5] S. Dutta Roy is Assoc. Professor in Electrical Eng. Dept., IIT Delhi, India. sumantra@ee.iitd.ac.in

[6]S. K. Saha is Professor in Mech. Engg. Dept., IIT Delhi, India. saha@mech.iitd.ac.in

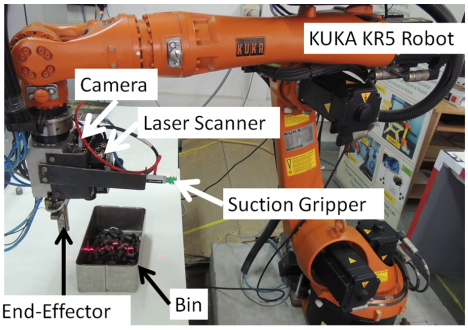


Fig. 1: Setup depicting randomly oriented pellets in bin. Sensors and vacuum gripper mounted on robotic manipulator.

A. Related Work

Methods reported for bin-picking for multi-layer configurations employ either specialised range sensors, structured lighting or Multi Flash Camera(MFCs) to build a 3-D representation of the bin contents. Lighting and image-based range sensing algorithm are sensitive to ambient conditions and are also misguided by specularity in metallic object [1]. Though work-arounds for such limitations exist, we focus on segmentation, pose estimation, pick-up and approachability analysis for Bin Picking, and emptying. RGB-D sensors like Kinect fail to capture enough detail for small dark objects like those in our system. Hence, we use a dedicated range sensor for depth measurements and an independent camera for image data. [2] Some recent approaches to this classical problem of bin-picking have seen approaches based on depth maps, with various assumptions to simplify the task at hand. Domae et al. [3] make a 4-DOF restricted movement assumption, and concentrate on picking objects without an explicit pose estimation, which may be important for pick-and-place tasks, and not just bin-emptying. 3-D segmentation using RANSAC-based methods ([2] and [4]) tend to be computationally expensive due to their iterative nature. Optimised RANSAC [5] is typically faster: the authors present an approach for picking large sparsely-placed cylindrical and T-pipes. The system is a bit slow for a fast bin-picking task. The paper reports a high number of false negatives, and is unsuitable for a case of a large number of similar objects with heavy occlusion. Similarly, iterative and computationally expensive approaches such as those of [6] may not be suited for a fast implementation. In approaches using active lighting, Liu et al. [7] report better timing and accuracy. Though methods relying purely on depth sensors have some limitations: in case of near perfect

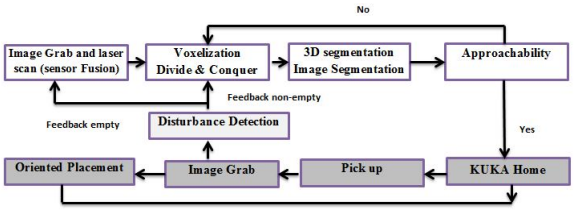
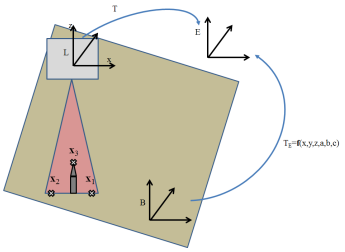


Fig. 2: Flow diagram depicting parallelism and visual feedback white boxes: PC control; Grey boxes: KUKA control.



(a)

(b) Schematic: laser scanner frame and robot tool frame

alignment of small identical objects, such methods fail to distinguish closely placed objects (as in Fig.6a). The problem with existing image-based approaches is for objects that lack edge contrast or have smooth curved surfaces. Further, effects like specularity create higher contrast than actual edges, triggering false detections. Liu et al. [7] discuss such limitations of image-based segmentation. Given a general situation, segmentation methods based on multi sensory input (grey-scale/RGB and depth data) [8] tend to work better than using any one sensor.

B. Our Contribution

The main contribution of this paper is the use of a low-resolution 3-D scan to guide a localized image search to optimally utilize the information from multiple sensors.
The paper is organized in the following way: Sec II presents the basic sensor calibration and fusion scheme. Sec III describes the segmentation and pose estimation of feature-less,symmetrical objects, uniform in color. This uses the sensor fusion and calibration of Sec II. Sec IV models approachability of an object,give its pose and position estimates in a bin. Sec IV also discusses time efficient techniques for emptying the bin. Sec V present results of segmentation, pose estimation, system repeatability as shown in Fig. 2. Sec VI presents a discussion and concludes paper.

II. Calibration and Sensor Fusion

The 3-D data obtained from translating the laser scanner(2-D scanner) and the image data has to be expressed in a common world coordinate frame (B). Earlier approaches [9], [10] involved directly obtaining the transformation of a laser scanner coordinate frame (L) from camera coordinate frame (C) . In [10], an 3-D scan from single view point is necessary and the application discussed by [9], was in mobile robot where high accuracy is not required and gave an error of 16 mm. This value is not suitable for our application, since the size of the objects is less than the error obtained. We had the advantage that the sensors were mounted on the end-effector of an industrial robot and we could rely on its positioning accuracy, thus the end-effector frame (E) was chosen as the reference frame to both L and C.

A. Transformation between Sensor Frames

Fig 3b shows diagram of the three coordinate frames. $\mathbf{T}_E$ was obtained by calculating the $[\mathbf{R}_E|\mathbf{P}_E]_E^B$ where $\mathbf{R}_E$ is the rotation matrix estimated from tool orientation and $\mathbf{P}_E$ is the translation matrix found using the tool position with respect to the world frame(B), using the tool position and orientation given by the robot controller.

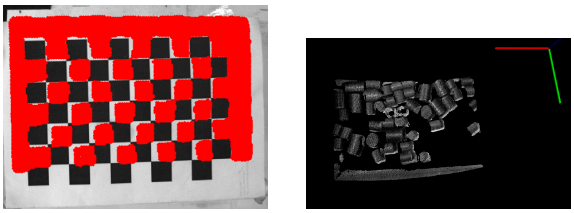
To find the transformation matrix $[\mathbf{T}]_L^E$, first a rough estimate ($[\mathbf{T}^{est}]_L^E$) was obtained. By positioning the laser scanner profile onto a pointed object three points were extracted from the profile in the XZ plane (as in Fig. 3b) of the laser sensor coordinate system. The robot end-effector was adjusted so that the points $[\mathbf{x}_1]_L$ and $[\mathbf{x}_2]_L$ have the same z coordinate. Using the geometrical constraints imposed an initial estimate for the transformation was obtained. For optimisation the 3-D coordinates of the grid cross section, were expressed in the tool frame (E) using the data from the robot's controller. These were then measured using the laser sensor by scanning the square calibration grid(Fig. 4a) at an exposure such that the black squares were not visible to the scanner. The crossectional point were also found in the image. An initial estimate of the transformation was done using Camera calibration toolbox [11]. Then, an estimate of the transformation ($[\mathbf{T}^{est}]_E^C$) from end-effector coordinate frame (E) to Camera coordinate frame (C) was obtained using [12]. The measured values of the crossectional points found using the camera and the laser sensor were compared against their estimate in the tool frame, to minimize the errors. An optimization method similar to the one used for robot's kinematic identification given in [13] was used to obtain rotation and translation parameters of both $\mathbf{T}^{est}$ to obtain the new matrix $\mathbf{T}^{opt}$.

B. Projection and Back Projection

The overall transformation of laser frame to camera frame is given by $[\mathbf{T}]_L^C = [\mathbf{T}]_E^C[\mathbf{T}]_L^E$. The projective map is established using the frame transformation $[\mathbf{T}]_L^C$ and the camera internal parameters. 3-D data obtained from the scanner was projected onto the image as visible in Fig. 4a, the final mean error was 1.2 mm. The inverse image of this map defines the back projection for selected pixels Fig. 4b.

III. Segmentation and Pose Estimation

Image based segmentation performance is affected due to variation in lustre, lighting and lack of contrast, whereas 3-D segmentation is completely impervious to such factors. Thus, we begin our segmentation process with the point cloud generated from the depth map. Later, the results



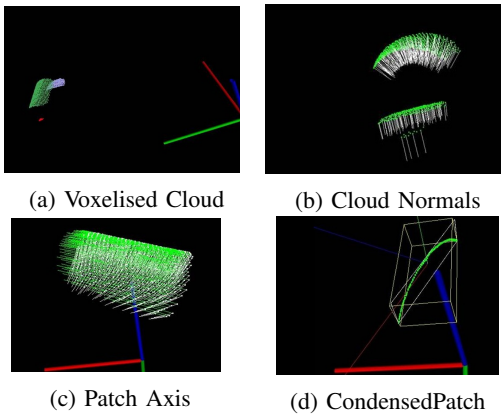
(a) Projection (b) Back-projection

Fig. 4: (a) Projection on calibration grid, (b) Back projection on objects in bin.

were processed by image based segmentation for missing information.

A. 3-D Segmentation

On receiving input data from the scanner, points corresponding to container walls and outside it were removed. The cloud was de-sampled to induce uniform sampling density and then statistical outliers were filtered as described in [14]. A uniform spatial density ensured better determination of geometric properties. Post conditioning, these points were further processed as per the following Algorithm 1.



(a) Voxelised Cloud (b) Cloud Normals

(c) Patch Axis (d) CondensedPatch

Fig. 5: 3-D segmentation (details Algortihm 1) a) Region growing in Voxel, b) Normals estimation, c) Axis calculated, d) Cylindrical cluster projected onto plane perpendicular to the axis.

The size of the object as estimated in the object frame was then compared with the actual dimension of the object. Three conditions may arise:

- If size is within acceptable deviation, then the object is qualified and the estimates are supplied for pickup.
- In case of heavy occlusion, the size of the patch was smaller than expected. Image was used for verification.
- When multiple pellets were aligned as shown in Fig. 6b the length and breadth exceeded the expected value. These were forwarded for image based refinement.

B. Image Segmentation

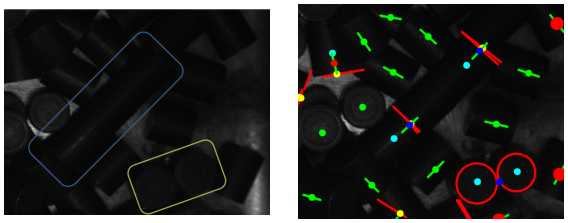
The projective map defined in II-B is used to find a correspondence between the point cloud and the image. This fusion of sensor data is further used to tune the results from

Algorithm 1: 3-D Segmentation

Input: Conditioned Point cloud data from Laser scanner .
Output: Surface Centroid and Approach Normal.
LOOP Process :

1: **Normal estimation:** Normals were estimated corresponding to each point in Voxel, as shown in Fig.5a and Fig. 5b
2: **Curvature estimation:** Variation in direction of normals of neighbours, was associated with curvature of each point.
3: **Region growing:** Patches were found using a region growing algorithm [14] which utilizes curvature value, direction of normals and euclidean distance between each point.
4: **Classification:** Each patch was classified to be belonging to specific shape (in our case plane or cylinder), based on mean curvature value.
5: **Centroid:** Mean of position of all points was the estimate for patch centroid.
6: **if** Surface is flat **then**
7: **Approach normal:** The average of direction of normals gave us, approach normal to the surface.
8: **else if** Surface is Cylindrical **then**
9: **Tangents and normals:** For each point, direction of maximum deviation of normals gave us the direction of the tangent as shown in Fig. 5c. Using the cross product of tangent and normal at each point the mean estimate for axis was computed.
10: **Cross-sectional plane:** All points on the cylinder patch were projected on the plane passing through axial center and with the axis as normal, Fig. 5d.
11: **Axis and approach normal:** Applying PCA to these points we got axis as the eigenvector with smallest eigenvalue and approach normal as the eigenvector with the second highest eigenvalue.
12: **end if**
13: **Object bounding box:** The length and breadth of the patch were also calculated by finding variation in projected lengths along basis vector of object frame.
14: **return** Centroid, approach normal and axis.

3-D segmentation while overcoming the usual limitations of image based segmentation as described in the Algorithm 2.



(a) Aligned Small Objects. (b) Segmentation Results

Fig. 6: Image-guided segmentation mitigates ambiguities in 3-D segmentation of multiple closely placed, similar objects.

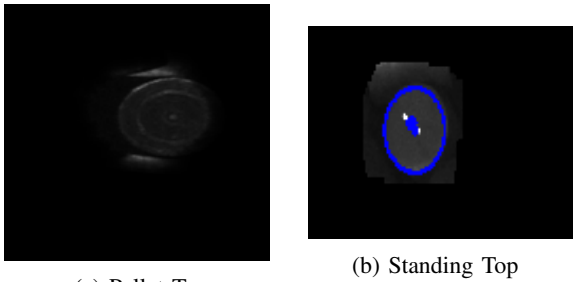
Algorithm 2: Image Segmentation

Input: Surface patch, type of surface, centroid, approach normal, axis, bounding box in object frame.

Output: Center.

LOOP Process :

1: **Projection:** Points from the surface patch are projected onto the image.
2: **Mask :** These points were used to create an image mask to define a ROI.
3: **Histogram equalisation:** To locally enhance the contrast histogram equalisation was applied to the ROI. Based on surface classification from 3-D :
4: **if** surface is flat **then**
5: **Circles:** Circles are fit to the equalized image and center is estimated. This provides the top of cylinder in the voxel, for results see Fig.7a and Fig.7b.
6: **else if** surface is cylindrical **then**
7: **Lines:** The axis calculated from 3-D data was projected onto the image. Lines were fit in the direction perpendicular the axis and center was triangulated using half it's projected length.
8: **end if**
9: **return** projected center



(a) Pellet Top

(b) Standing Top

Fig. 7: Using cluster extracted from 3-D a) shows defined ROI and b) shows segmentation result for standing pellet.

As opposed to when applied on complete image, Histogram equalization of the localized patch, greatly increases the contrast. This provided the requisite detail for successful image segmentation and also imbued immunity to variance in lighting conditions. By looking for specific features in localised region of high contrast, detection of false edges due to specularity is circumvented and consistent results were achieved. The center obtained from the image was back projected onto the cloud and using the estimates of Orientation and Approach Normal acquired from 3-D data, pellet pose and position were calculated. The key insight is to enhance contrast by localised processing and using the surface type and object orientation to guide the image search. Our algorithm has shown reliable performance and is resilient to change in ambient lighting.

IV. Setup

Basler camera of resolution 2448 x 2048 for image and Micro Epsilon high accuracy scanner for 3-D data provide inputs for the system. Both sensors were mounted on the tool tip in 'Eye in hand' Configuration. The algorithm ran on 2.8Ghz Intel Xeon CPU. With a single thread dedicated to all activities associated with segmentation. Three other threads were spawned. The First was for communication with KUKA, Second for inter process communication for independent operation of KUKA and segmentation threads. The last thread was for visualizing cloud data at time of pick up. We now briefly discuss key aspect in implementation of the algorithm in our set up, Fig. 1.

A. Segmentation, Pose Estimation and Graspability

The cylindrical pellets were found and orientation was estimated by the proposed algorithm. Pellets used were uniformly black in color and cylindrical in shape. They are 14-17 mm high with a radius of 6 mm deviating upto 0.5mm. These were piled in a stainless steel container, as much as 8 cm high. By varying these parameters, the implementation could be extended to cylindrical objects of different sizes. Graspability was ascertained by comparing patch dimensions in object frame with the gripper diameter. If qualified the surface centroid and approach normal information were translated to KUKA Robot in the form of X,Y,Z and A,B,C. For this the Z axis of the gripper rod was aligned with the estimated approach normal and X axis with the estimated axis of the object.

B. Approachability

In order to have a reliable system for unloading of the bin completely, it becomes necessary to plan an optimal path considering the approachability of the pellet. Due to arbitrary and complex orientation of pellets, collision between the gripper and container walls is likely. Thus, before attempting pick up we ensured that the required orientation of the gripper was feasible. We define an analytical approach, inspired from [15] where cylinder to cylinder interaction was modelled. Our approach is different as we model cylinder and plane collision by approximating the gripper rod with a cylinder and the walls of a container by finite planes. The cylinder was assumed to be oriented with axis aligned along the surface normal of the object, given by $(n_x, n_y, n_z) \epsilon \mathbb{R}^3$ and with the center of cylinder's face coincident with patch surface centroid, given by $(p_x, p_y, p_z) \epsilon \mathbb{R}^3$ .It was then analysed for proximity to planes.

The analysis begins by projecting the geometric model of the system onto the xy plane which is parallel to the table. We then found the planes in direction of orientation of the cylinder that could possibly suffer collision.Let $f\colon \mathbb{R}^2 \rightarrow \{(AB, BC)\,(BC, CD)\,, (CD, DA)\,, (DA, AB)\}$ where AB, BC, CD, DA represent planes projected as lines in the XY plane as shown in Fig. 8a.

$$f\left(n_x, n_y\right) = \begin{cases} (AB, BC) & 90 > \tan^{-1}\left(\frac{n_y}{n_x}\right) \geqslant 0 \\ (BC, CD) & 180 > \tan^{-1}\left(\frac{n_y}{n_x}\right) \geqslant 90 \\ (CD, DA) & 270 > \tan^{-1}\left(\frac{n_y}{n_x}\right) \geqslant 180 \\ (DA, AB) & 360 > \tan^{-1}\left(\frac{n_y}{n_x}\right) \geqslant 270 \end{cases}$$

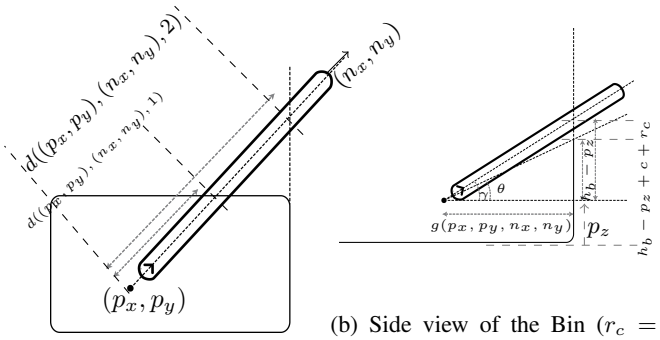
To calculate distance of both walls from the pick point along pick up orientation, if $f(n_x, n_y) = (AB, BC)$ then $f(n_x, n_y)(1) = AB$. Now let $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$, where $d((p_x, p_y), (n_x, n_y), i) =$ distance of point (x,y) from along direction (n_x, n_y) from line $f(n_x, n_y)(i)$. Then the distance pick up point from the wall closest along the cylinder axis is given by $g\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, where

$$g((p_x, p_y), (n_x, n_y)) = min_i \; d((p_x, p_y), (n_x, n_y), i) \quad (1)$$

For the following steps we restrict the discussion to the plane defined by the cylinder axis and the projected normal given by the line $(p_x, p_y, p_z) + t(n_x, n_y, 0)$. Let the overall angular clearance required = γ, known height of the bin be h_b and $h\colon \mathbb{R}^2 \to \mathbb{R}$, where $h(z) = h_b - p_z$. Then, $\tan(\gamma) = (h(z) + r/\cos(\gamma))/g(p_x, p_y, n_x, n_y)$ and $\sin(\gamma) * g(p_x, p_y, n_x, n_y) - \cos(\gamma)(h(z)) = r$. Let $\tan(\alpha) = g(p_x, p_y, n_x, n_y)/h(z)$, then

$$\gamma = sin^{-1}\left(r/\sqrt{(h(z))^2 + g(p_x, p_y, n_x, n_y)^2}\right) + \alpha \quad (2)$$

The angular clearance offered as per current cylinder configuration $= \theta_{(n_x, n_y, n_z)} = \tan^{-1}\left(n_z/\sqrt{n_x^2 + n_y^2}\right)$. The path is collision free if $\theta > \gamma$ else if $\theta < \gamma$ but $\gamma - \theta < 15\,\text{deg}$ then the approach normal was adjusted to re-route the path and approach without collision. If the adjustment required was beyond 15 deg the suction gripper fails in pickup thus those cases were skipped. Similar approach is employed for sides of the cylinder. Choosing the line of interest by shifting the cylinder axis sideways, by the radius of the cylinder. The above algorithm ensures the gripper circumference do not collide with the container wall. The planar face of cylinder doesn't collide with the wall by ensuring that the minimum distance of any wall from the face centroid is greater than the projected radius of the cylinder.



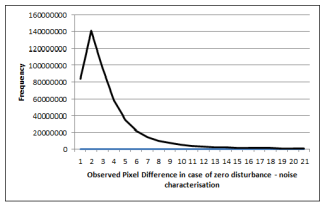
(a) Top view of the Bin

(b) Side view of the Bin ($r_c = r/\cos\gamma$ and c is clearance from bin to cylinder)

Fig. 8: (a)XY Plane,(b) Cylinder-axis and Z-axis plane

C. Disturbance Detection

During pick up, there were other sources of disturbance besides collision with the container, which were hard to model. These were caused due to object-object interaction or object and gripper interaction. To cater to such disturbances without initiating a fresh scan an image is taken of the container from the same perspective, after each pick up. This image was compared with the previous image. Pixel by pixel difference was calculated in both images, as shown in Fig. 9b. If the pixel difference was found to be above the noise threshold decided by noise frequency given in Fig. 9a the change was considered to be caused due to movement or disturbance in the scene. Range data corresponding to this pixel was deleted using the method of backward projection established in II-B. This ensured that these points were not used in the segmentation process of the next pellet. Thus, the disturbed pellets were successfully ignored while circumventing the need of a rescan.



(a) Noise Threshold

(b) Disturbance Detection

Fig. 9: (a) Noise threshold estimated from repeated trials, (b) pixels highlighting disturbance detected.

D. Oriented placing and emptying of Bin

The objects were placed in uniform pose using the estimate of orientation and keeping in consideration the approximations made to make the approach feasible. We considered that a cycle begins from a scan and terminates at the requirement of another scan. Multiple pellets were picked up in a cycle. Some pellets were not picked up to avoid collision between gripper and the container. such an attempt may result in misalignment of all pellets. If the number of pellets picked up in a given cycle, was very less as compared to the total number of pellets segmented in the cycle, then the bin is shaken. The condition is imposed as the time cost per pellet for conducting a scan increase as the number of pellets picked in a cycle decreases.

V. Experiments and Results

Pellets were placed in the aluminium bin in multilayer configuration in random orientation. Cases of severe occlusion were tested by piling up pellets in heap. Average Scan time was 8 seconds. Though Scan timing would vary based on equipment and robot configuration, it puts in perspective the time that is saved by reducing the number of scans required for emptying the bin. Complete timing for pickup process has not been discussed as it shall be highly subjected to tool configuration, robot model and operating speeds.

A. Segmentation and Pose Estimation

The segmentation algorithm was tested on a set of ten scans and images. Each containing 125 pellets of which 70 to 80 would typically be visible. Objects approximately

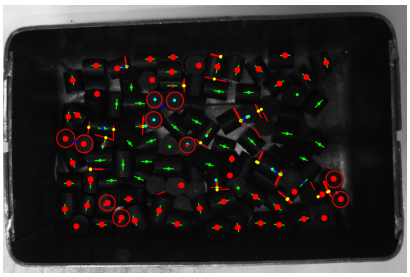


Fig. 10: Green color represents of surfaces as classified by 3-D analysis. Red colour is associated when results from 3-D segmentation are non-conclusive.

40% exposed, were considered as visible. Our algorithm successfully determined the position and orientation of 70% of the visible objects. While dedicatedly running on single thread the process took 300 ms to isolate and find the attributes for a single pellet. High detection rate in a single scan allows locating multiple pellets in parallel to the robot motion as shown in Fig. 10. This eliminated processing delays and also reduced on requirement of multiple scans for consecutive pickups.

B. System Repeatability

When running the whole algorithm for emptying of the bin. Success rated at 93% in 1200 pick ups, the failure cases include misclassification between planar and curved surface(2% cases when pickup is successful but orientation is missed). 2% were associated with segmentation failure such as the, centroid estimate of the exposed portion is too close to the boundary of the object. 3% of misses were attributed to disturbance of object while pick-up process due to gripper moment and suction failure. Please note the objects were in a pile and lack ground support. For most oriented object even slight disturbance caused trouble some shift.

VI. DISCUSSION AND CONCLUSIONS

We showcase an alternative approach to multi-sensor based segmentation techniques. The proposed segmentation and pose determination algorithm builds on information generated by the 3-D segmentation process to refine the search space for the image segmentation. Though depending upon RGB data the algorithm shows immunity to illumination variance by virtue of localised image processing. The approach counters the limitations posed by effects like specularity, axial alignment, facial alignment and low contrast. This introduces a new class of algorithms which would look for specific information from different sensory inputs. Utilising the strength of the individual sensors to improve overall performance of the system. Purely geometric methods like RANSAM and RANSAC tend to be time consuming and are limited by the sensor resolution. Supporting these techniques with image segmentation would help speed up the process and increase accuracy. Also our approach is deterministic and would ideally be faster than iterative counterparts.

The integrated disturbance detection module decreases the amortized pickup cost per pellet by saving on the time required for repetitive scan and data conditioning. The adaptation of the system in part or completeness would help improve performance in existing setups. The overall approach along with studying the approachability offers a comprehensive solution for emptying the bin and is suited to industrial implementations owing to its performance and robustness. Further, we are planning to generalize our algorithm for objects of different primitive shapes using Deep Learning based approach as in, [16]

ACKNOWLEDGMENT

A. Kumar, S. Sen, A. A. Hayat and R. P. Joshi helped in experiments and calibration of the set up.

REFERENCES

[1] M. Gupta, Q. Yin, and S. Nayar, "Structured Light in Sunlight," in *IEEE International Conference on Computer Vision (ICCV)*, Dec 2013.

[2] H.-Y. Kuo, H.-R. Su, S.-H. Lai, and C.-C. Wu, "3d object detection and pose estimation from depth image for robotic bin picking," in *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1264–1269.

[3] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai, "Fast grasp ability evaluation on single depth maps for bin picking with general grippers," in *in Robotics and Automation (ICRA), 2014 IEEE International Conference on*, Oct 2014, pp. 1997–2004.

[4] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3d geometry matching for grasping of known objects in cluttered scenes," *The International Journal of Robotics Research*, p. 0278364911436019, 2012.

[5] N. Matthias, D. David, H. Dirk, S. Joerg, B. Alexander, L. Jun, and B. Reinhard, "Mobile bin picking with an anthropomorphic service robot," in *in Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 2327 – 2334.

[6] D. Buchholz, M. Futterlieb, S. Winkelbach, and F. M. Wahl, "Efficient bin-picking and grasp planning based on depth data," in *in Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 3245–3250.

[7] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. K. Marks, and R. Chellappa., "Fast object localization and pose estimation in heavy clutter for robotic bin picking," *Int. J. of Robotics Research*, vol. 31, no. 8, pp. 951–973, 2012.

[8] R. B. Rusu, M. Beetz, Z. C. Marton, N. Blodow, and M. Dolha, "Towards 3d point cloud based object maps for household environments," *Robotics and Autonomous Systems Journal (Special Issue on Semantic Knowledge)*, 2008 2008.

[9] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3, Sept 2004, pp. 2301–2306.

[10] R. Unnikrishnan and M. Hebert, "Fast extrinsic calibration of a laser rangefinder to a camera," *CMU Twechnical Report*, 2005.

[11] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.

[12] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.

[13] W. Khalil and E. Dombre, *Modeling, Identification and Control of Robots*, 3rd ed. Bristol, PA, USA: Taylor & Francis, Inc., 2002.

[14] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," Ph.D. dissertation, TU Munich, Munich, Germany, 2009.

[15] R. Chittawadigi and S. Saha, "An analytical method to detect collision between cylinders using dual number algebra," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 5353–5358.

[16] M. Schwarz, H. Schulz, and S. Behnke, "Rgb-d object recognition and pose estimation based on pre-trained convolutional neural network features," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1329–1335.