Efficient Object Isolation in Complex Environment Using Manipulation Primitive on a Vision Based Mobile 6DOF Robotic Arm

Quah Jit Shen[1], Muhammad Fahmi[1], Khalil Azha Mohd Annuar[2]

1) CERIA Research Lab, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

2) Faculty of Engineering Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

***Abstract*– This paper explores the idea of manipulation aided-perception in the context of isolating an object of interest from other small objects of varying degree of clusterization in order to obtain high quality training images. The robot utilizes a novel algorithm to plot out the position for each noise objects and its destined position as well as its trajectory and then utilizes manipulation primitives (pushing motion) to move said object along the planned trajectory. The method was demonstrated using V-rep simulation software which simulated a Kuka YouBot fitted with a camera on the gripper. We evaluated our approach by simulating the robot manipulators in an experiment which successfully isolate the object of interest from noise objects with at a rate of 77.46% at an average of 0.56 manipulations per object compared to others at 1.76 manipulations subsequently speeding up the time taken for manipulation from 12.58 minutes to 2.6 minutes however suffers from a tradeoff in terms of accuracy when comparing the similar works to our proposed method.**

Index Term-- **Trajectory Generation, Sorting, Object Manipulation, Object Isolation**

1. INTRODUCTION

This research proposes a novel algorithm which utilizes manipulation primitives to perform object isolation by manipulate the complex scene into an idea scene for the purpose of obtaining a better-quality training image. Our proposed method is capable of isolating an object from a complex scene of varying degrees of isolation simulating real unstructured environment. In contrast to purely image processing approaches our method can effectively remove noises and solve occlusions of objects which may otherwise lead to incorrect perception [1,2] and ultimately failure in future operations [3,7,8]. We demonstrated the effectiveness of our single manipulation approach using a 6 DOF KUKA YouBot to isolate bricks of a specific colour away from the single brick of interest with another colour.

Object isolation in essence is the problem of removing a clutter or unwanted objects from the working area. Traditionally the use of image processing is often employed to phase out or counter react against the noises that occur from the environment, such as occlusion or the object of interest blending with the background however solving an image processing problem through the help of environment manipulation method is not a new thing in the field of research nor in nature as it is a common practice to displace objects which are unrelated and that is in the way to better perceive the object of interest [1-14]. Thus, this becomes a motivating factor for this research in robotics manipulation of environment to help its perception of understanding of the environment. The manipulation of unwanted objects from a scene can be easily accomplished using manipulation primitives such as pushing, pulling, picking and dropping. However, while the manipulation itself may not be an issue, the planning for manipulation motion/strategy, motion variables/ parameters i.e. initial position, final position and path needs to be identified and may prove to be a challenge due to its complexity in regard to the environment, surroundings as well as the objects characteristics and behaviour.

2. RELATED WORK

Object isolation is a niche field of study which stems from a larger group of study known as pile manipulation/ interaction [4-8]. Within this field, the closest work on isolation is known as object singulation where objects which exist in a pile are relocated and/or reposition for future actions. Commonly, the method of manipulation primitives employed was pick and place due to the requirements of precision in the act of relocating/repositioning the objects from a pile is assisted by perturbing pushes as a means to loosen/separate the objects that were in a pile to make it easier for grasping to occur [4,8]. The research conducted by Chang et al [8] done successfully manipulated 70% of the items from the original pile with an error rate for the adaptive model combining both perturbing pushes for repositioning followed by grasping was reduced to 2 errors in grasp out of 46 grasping attempts one for an empty grasp and one for a lost item in comparison to when only grasping motion is used without an initial perturbing push to reposition an item which had an error rate of 16 error in grasping out of 52 grasp attempts. This result highlights the positive effects in the usage of motion to improve perception. However, there were limitations in terms of the efficiency of the pushes in introducing large enough motion to split the target state into multiple spatial units in terms of number of pushes required to successfully allow the grasp attempt as well as the time taken.

To solve the issue raised in the previous research which uses small pushes to release an object from a pile to successfully employ a grasp for object singulation, a different method of manipulation primitive can be employed as seen by the research done by Katz et al [7] who research aims was to clear a pile of unknown objects using interactive perception. In this research, they manage solve the issue of efficiency of pushes by removing the usage of pushes and introducing a poking motion along the z-axis on the object before employing picking and placing to remove the object from the scene. Using this method, they managed to reduce the number of action for successful grasp from 6.6 manipulation per grasp from previous research [8] to 2 manipulations per grasp with a total operation time for a full manipulation averaging out to be 34 seconds for a full operation. However, there is an issue with using this method of manipulation primitive to improve perception and that comes in the form of the fact that poking may fail to move the object enough or cause a significant enough disturbance to the environment which causes matching to fail and subsequently forcing the robot to poke again increasing the number of manipulations per object. Next, like the previous research, there still exist problems pertaining to the robot not being able to grasp an object from the scene and remove it given that the object is too big or too small for the hand to grasp.

From the issues raised by the previous research [7], a combination method was devised to deal with the issue of clusterization and piling up which was previously inadequately dealt with by Gupta et al [5,6]. In this research, 3 manipulation primitives were introduced which were pick and drop for relocating objects, tumbling for decluttering a pile and spreading to declutter a region. By having the additional method of tumbling which is an extension of the pushing method, the research could solve the previous problem of poking not being able to move the object enough for a successful manipulation and this is represented by the system being able to reduce the number of successful actions per manipulation from the previous 2 manipulations per grasping in [7] to 1.76 manipulation per grasping [6] with an accuracy rate of close to 100%. It however does not solve the issue of inability to grasp objects which may be bigger than the gripper similar to the previous researches [7,8].

A different method to solve the problem of obtaining higher quality data would be instead of altering the complex scene itself as stated in previous research [5-8], we instead could manipulate the orientation of the camera to an angle which it is more capable of obtaining higher quality data [9,10,13,14]. A research done by C. Dornhege and A. Kleiner [14] uses point cloud technology to map out an image and identify what it calls a void cell which is an opening in a structure which indicates that there is a space confine within said structure and thus by looking out for these void spaces and scanning the area indicated by this void space, it is possible to better perceive the environment without manipulating the environment or employing more complex image processing techniques. In this method, for a single scene the robot requires an average of 6 manipulations/scans with an average time of 3.5s per manipulation/scan to fully perceive the environment that was presented to it. A problem with this method however is that it is only capable of obtaining a single high-quality image of an object of interest through its multiple manipulations because while it is capable of obtaining data if an object is occluded by an object with a certain distance away from it, it is not able to discern information if the object that is occluding is in direct contact with the object of interest thus potentially hiding away important information such as actual size or shape of the object of interest when multiple noise objects are piled on top of it.

While the methods mentioned above all manages to solve a lot of the problems involved with environment manipulation for the purpose of increasing perception, the methods employed from [4-8] are mainly used for sorting and while it is equally applicable to isolation, it is simply not as efficient simply because in the act of sorting, all blocks are important thus the system have to delicately take it’s time to rearrange the arrangement of objects in the vicinity whereas in the case of isolation, there is only 1 object of importance and the goal of the program is to remove the rest of the unrelated objects hence while it is possible to utilize delicate object manipulation methods, it is simply inefficient when considering that we can remove multiple object at once directly using tools [15] or indirectly using manipulation casualties and this can be seen in [8] where selecting and subsequently isolating each and every individual objects takes a long time (6.6 manipulations for a single object) when again taking into consideration of the fact that there is only one object that is of interest . Our method proposes the use of a single manipulation primitive (pushing movement) to isolate an object efficiently from a cluttered/complexed environment exploiting the effects of manipulation casualties to remove more than one object per manipulation hence reducing the overall time/manipulations required to fully isolate the object.

3. ALGORITHM

We consider the problem of isolation of objects scattered or piled n on the effective workspace based on the colour of the object where green marks and object of interest and blue marks a noise object. The algorithm comprises of 4 repeated steps: 1) Object segmentation, 2) Case base reasoning action sequence 3) planning of arm motion for the subsequent action and 4) plan execution and control whose process can be seen in Fig. 2 and outcome can be seen in Figure 1. This paper focuses on the case-based reasoning algorithm used to perform the isolation as well as the usage of manipulation primitives (pushing motion) as a form of effective isolation. Our algorithm choses its actions and then isolates the objects for the purpose of improving visibility of the object of interest from all 360° angles while still on the workspace from a cluttered/ unstructured/ complex environment as described below:

1) **Segmentation into category:** Objects are segmented into 2 distinct categories which are object of interest

and noise objects through the colour property where green refers to the object of interest and blue is the noise object.

2) **Manipulation of region:** To perturb the region, the research proposes using a simple manipulation primitive (pushing motion) which is applicable to a variety of other robots with or without a gripper end effector and at the same time generic enough to be able to applied on a variety of objects. The exact path planning for the manipulation was determined based on two conditions which are:

- *Y Coordinate of the Detected Noise object centroid is within ∓0.05units from the centroid of the object of interest:* In this situation, manipulator positions itself parallel to the noise object centroid along the y axis with the initial coordinate at $y_{border-min}$ and final coordinate at $y_{border-max}$ as defined in Fig. 3
- *Y Coordinate of the Detected Noise object centroid is not within ∓0.05units from the centroid of the object of interest:* In this situation, manipulator positions itself parallel to the noise object centroid along the x axis with the initial coordinate at $x_{border-min}$ and final coordinate at $x_{border-max}$ as defined in Fig. 3

3) **Repeat** until the workspace is cleared of all noise object leaving the object of interest totally isolated.

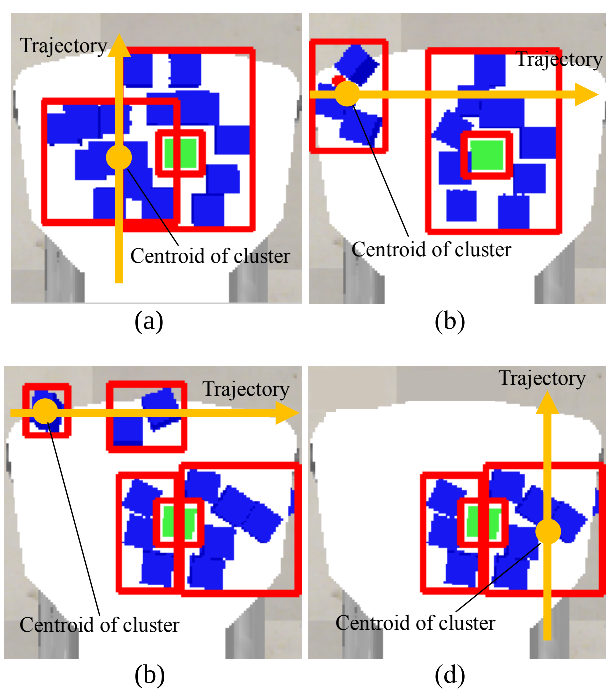


Figure 1: Example of how the algorithm function process by process (a) the centroid of the cluster is determine to be on the same level to the object of interest thus Y axis manipulation is actuated (b) cluster centroid is far from center thus X axis manipulation is actuated (c) outermost object is designated as a target and X axis manipulation is actuated (d) Outermost cluster is designated and because centroid is align with the object of interest Y axis manipulation is actuated

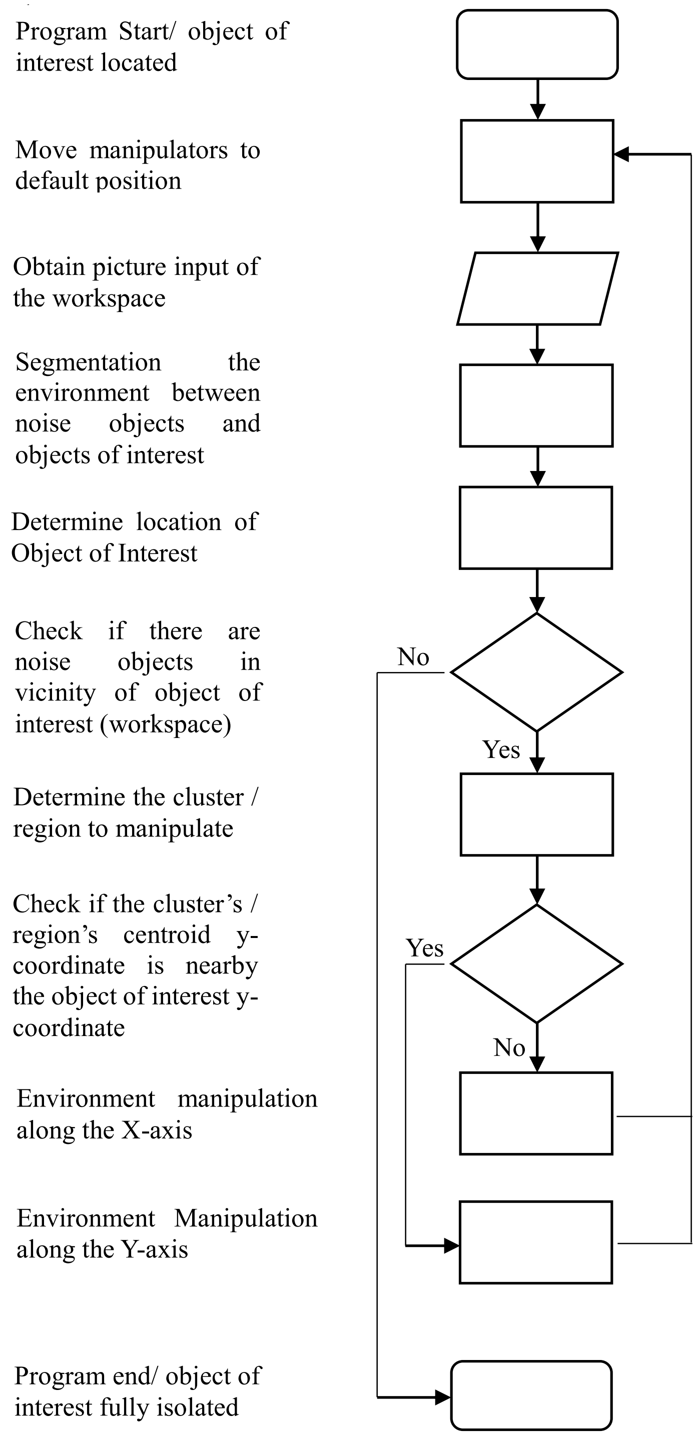


Fig. 2. Flowchart showing the decision-making process for the system in isolating the noise objects from the object of interest

By utilizing only pushing as the mode of manipulation for the purpose of isolation, we can make use of the indirect manipulation to remove other noise objects that are cluttered together out of the workspace more effectively than mere pick and place thus reducing the expected total manipulation per object and subsequently decreasing the overall time for task completion.

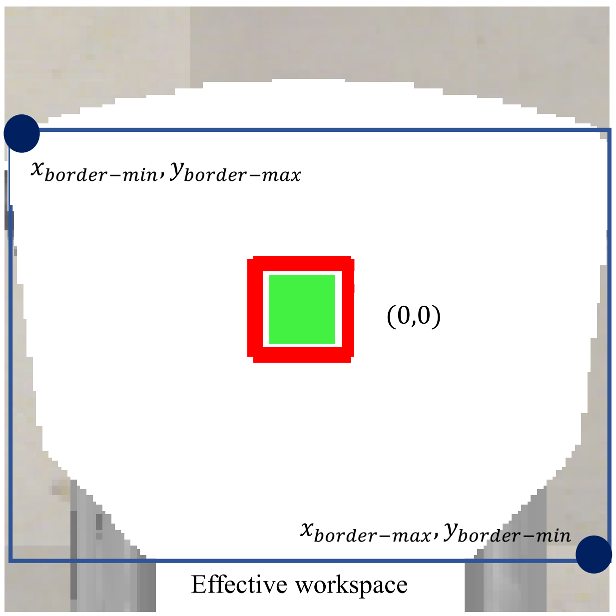


Fig. 3. Effective workspace of the research experimental setup

4. IMPLEMENTATION

We implemented and experimentally evaluate the single manipulation method and algorithm using the Kuka Youbot which is a 6 Degrease of Freedom Robotic arm with a two-finger gripper as its end effector all of which is attached to a mobile platform. The Kuka Youbot is then fitted with four cameras which has resolution of 128 x 128 pixels and is subsequently mounted onlooking each angle towards the middle of the workspace as seen in Fig. 4. Only the camera mounted directly above the workspace is used for image processing whereas the other 3 cameras are used as a means to evaluate the quality of training image that will be captured in any given circumstances. All of this was done using V-Rep simulation software linked with Python X,Y for the programming bit. All image processing methods was done utilizing OpenCV library which was installed into the Python X,Y compiler.

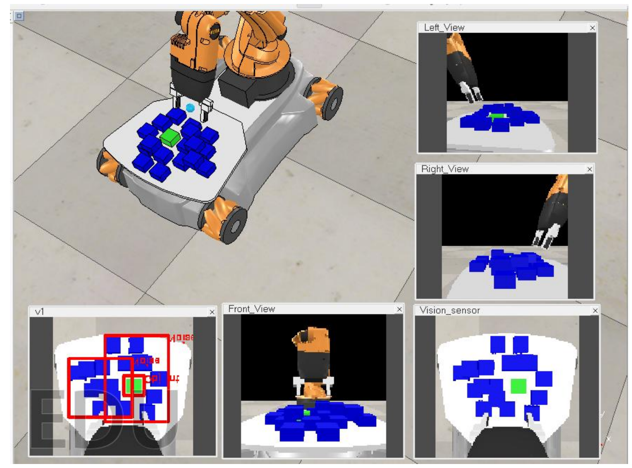
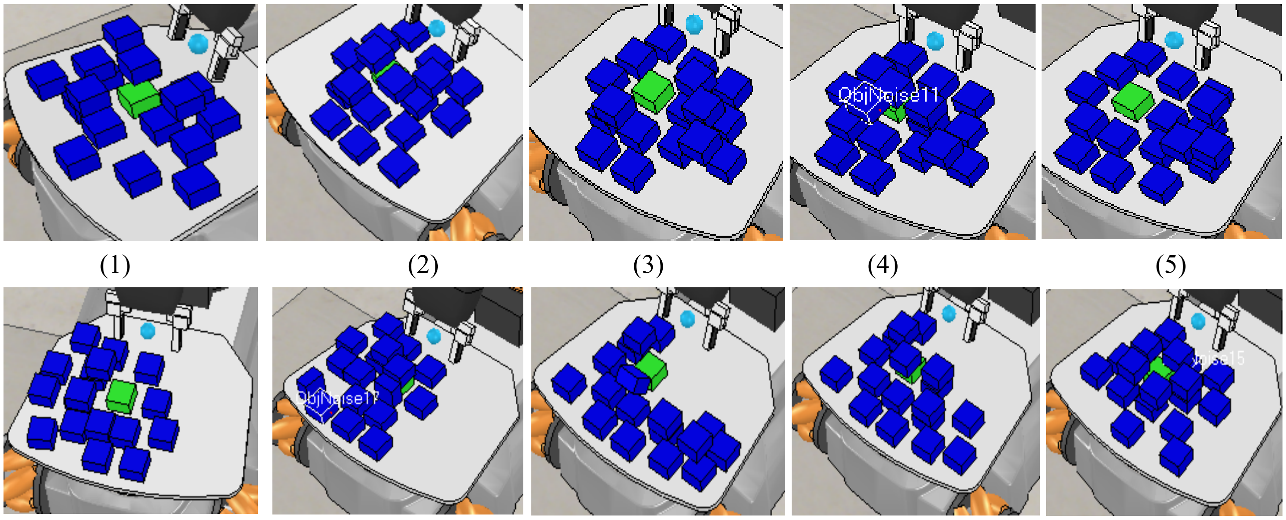


Fig. 4. Full Experimental Setup: Kuka robot arm mounted on a platform looking at the simulated Duplo bricks in V-Rep simulations on the effective workspace. The task is to remove the noise object leaving only the object of interest in the effective workspace

In this paper, we apply the isolation algorithm and method on 18 simulated 2x2 Duplo Bricks with the dimensions of 32mm x 32mm x 19mm. A number of Duplo blocks are scattered within the effective workspace along with a single object of interest. For the purpose of effective object position awareness, the effective workspace was mapped in a cartesian plane which in turn is valued corresponding to the kuka robot arm coordinates where coordinate 0,0 is located at the center of the effective workspace corresponding with the 0,0 position for the kuka youbot, $x_{border-min}$ and $x_{border-max}$ is the x coordinate for the left most extension of the robot arm and right most extension of the robot arm respectively whereas $y_{border-min}$ and $y_{border-max}$ is the y coordinate for the bottom most extension of the robot arm and top most extension of the robot arm respectively as seen in Figure 1 and Fig. 3. The environment itself was randomly scattered as seen in Fig. 5 however the blocks were still scattered in a way that shows possible different variations in situations where situation 1-3 are scattered radially with its highest density in the middle where the object of interest is and decreasing as it reaches the outer areas of the circumference with maximum stacking height of 2 blocks. Case 4-5 on the other hand was scattered such that there were more blocks on the right of the object of interest, Case 6-7 is scatter where there are more blocks on the left side of the object of interest whereas 8-9 is scattered such that there are more blocks at the front of the object of interest and last but not least case 10 was scattered such that the blocks piled up to stacks of 3 heavily concentrated at the middle above the object of interest

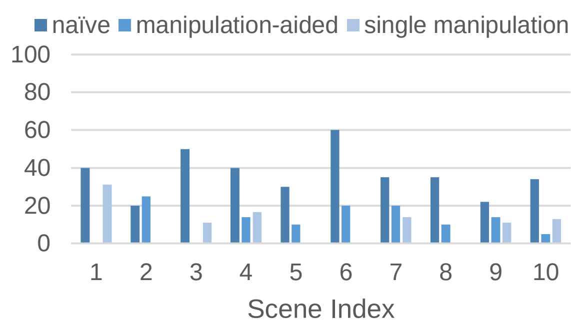
The clustered environment was meant to simulated different variations of a cluttered environment common in reality with the given caveats. The idea behind it is that often times there are numerous objects which are left around either intentionally or non-intentionally which is obstructing a good picture of the object of interest from being taken.



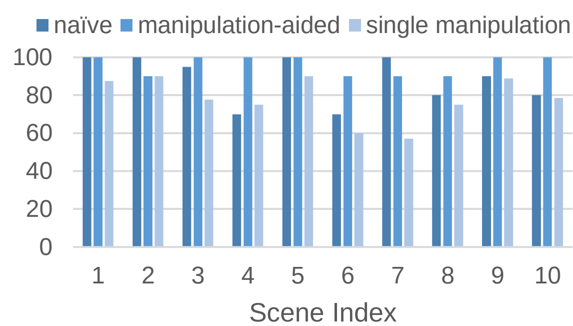
(1) (2) (3) (4) (5)

(6) (7) (8) (9) (10)

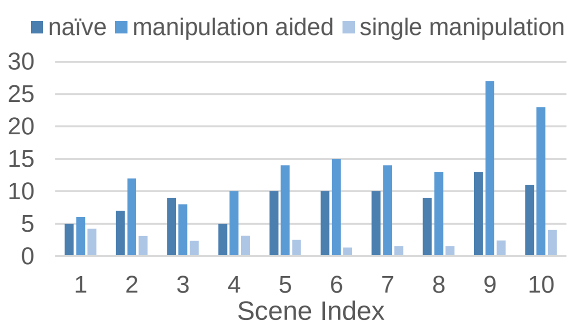
Fig. 5. The performance of the object isolation algorithm on various matrices was done for the ten configurations shown in the figure. The results are reported for a single trial for each configuration in Fig. 6



(a)



(b)



(c)

Fig. 6. Comparison of performances of naïve and Gupta et al sorting method against our single manipulation method on various metrics on ten configurations as shown in Fig. 5. The results are reported for a single trial for each configuration and the percentage are with respect to the number of objects. (a) Percentage of Failure, (b) Percentage of successfully sorted objects, (c) Time taken (minutes)

5. RESULT

We compared our single manipulation primitive (pushing motion) approaches to a naïve approach and Gupta et al's manipulation-aided approach [5] with the caveat that our single manipulation method is employed to isolate thus we time taken to complete the operation as seen in Fig. 6 (c) is the time taken to completely isolate the object from its environment whereas for Gupta and naïve it is the time taken to successfully sort all the Duplo blocks into its respective groupings albeit with similar number of blocks (18) and almost similar pile configuration. Given that caveat, it can be seen that the performance in terms of total time taken to complete the task is significantly less at an average of 2.6 minutes compared to 14.18 minutes for naïve and 12.58 minutes for Gupta. The huge difference in time variance can once again be due to the caveat mention prior because sorting requires precise object to object manipulation thus accidental casualty of manipulation is avoided at all cause to ensure objects that are already sorted do not accidentally get displaced. On the other hand, for isolation, the position of noise objects and its subsequent positioning is less of a cause for worry because the ultimate goal is to remove them from the scene thus the casualty of manipulation actually aids in the process of noise removal effectively cutting down the manipulation time.

One of the matrices of our evaluation is successful grasp. We categories manipulation failures as the following:

- **Empty Manipulation:** A manipulation is actuated and the gripper moves to and successfully comes in contact with the noise object but fails to remove the/a object from the scene
- **Lost Object:** A manipulation is actuated and the gripper moves to the noise object however it fails to come in contact with the object due to it previously being pushed out of range of the end effector while still within the scene. In this situation, the object is removed from the scene manually.
- **Accidental removal:** When the object of interest is accidentally removed out of the scene due to the effects of casualty of manipulation. In this situation, the program is allowed to continue with the same rules until the workspace is cleared.

Using the categorization above, the results in Fig. 6 (a) which is the percentage of failure which is calculated using the formula

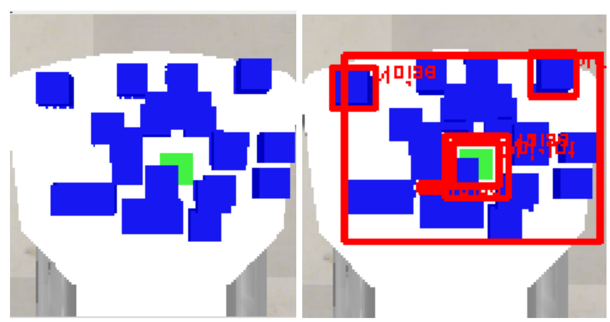
$$\%_{failure} = \frac{N_{empty\,grasp}}{N_{total\,grasp}} \quad (1)$$

It can be seen that our single manipulation method boosted a lower number of % error at 9.75% (with minimum 0 failures) compared to naïve which averaged out at 37% failures (with at least 20% failures) and Gupta et al at 12% failure (with minimum 0 failures). This reduction in rate of failure may be due to the fact that naïve and Gupta considers double grasping which is when multiple objects are manipulated simultaneously (casualty of manipulation) as an error whereas we use this as a feature to boost the efficiency of our method.

Fig. 6 (b) on the other hand calculated the percentage of successfully sorted objects in the case of naïve and Gupta and isolation in our case which was calculated using the following formula

$$\%_{sucess} = \frac{N_{total\,grasp} - N_{Lost\,object} - N_{accident}}{N_{total\,grasp}} \quad (2)$$

It can be seen that our singular manipulation method had a success rate of 77.46% (minimum 57% success rate) which is a significant drop compared to naïve which had a rate of 90.3% (minimum 73.3%) and Gupta et al with 97.8% (minimum 93.3%). This significant drop in success rate when compared to naïve and Gupta may be due to the fact that we have an extra category of manipulation which we consider an error which does not occur when doing sorting and that error is the accidental removal error which is where the object of interest is accidentally removed from the scene due to the effects of casualty of manipulation. In the experiment done, this effect occurred twice which was during (2) and (7) for the 11^{th} and 2^{nd} manipulation order respectively. In both cases, the object was interest was accidentally removed from the scene directly due to its interactions with the kuka arm while moving along the y axis rather than it being pushed out due to its interactions with objects and thus we surmise that it may be due to the programming for the orientation of the robot manipulator which was located too low of an angle compared to a desired 90° . Accidental removal of the object of interest from the workspace can be attributed to the quality of image processing which missed characterized a cluster of loosely positioned objects as a singular large object as seen in Fig. 7



(a) (b)

Fig. 7. (a) Actual Scene as seen by the robot, (b) Scene as classified by the robot's image processing algorithm

Due to the missed characterization, proper actions to countermeasure the closeness of objects are bypassed causing the object of interest to be remove accidentally via manipulation.

Last but not least in the terms of average manipulation per object as well as time taken per manipulation, it can be obtained that the average manipulation per item as well as average time can be obtained using the following formulas

$$Manipilation\ Per\ Object = \frac{\sum_{n=1,2,3} \frac{N_{actions_n}}{N_{objects}}}{n} \quad (3)$$

$$Average\ Time = \frac{\sum_{n=1,2,3} \frac{T_{time\ taken(s)n}}{N_{actions\,n}}}{n} \quad \textbf{(4)}$$

Where n is referring to the experiment number. Using this formula, the manipulation per object was found to be 0.56 manipulations per object which gives us an average time of 17.11s which is a significant increase compared to naïve which requires an average of 1.34 manipulations and 33.5s per successful manipulated Duplo block and Gupta which required 1.78 manipulations and 44.7s per successful manipulation.

6. SUMMARY

This paper explored manipulation aided perception in the context of isolating small objects (Duplo Bricks) off the kuka Youbot workspace. We presented an algorithm that combines perception and manipulation to efficiently and quickly to isolate unwanted noise objects from the object of interest using a single manipulation method. Preliminary simulation experiments result on the kuka Youbot simulated on Vrep successfully reduced the total number of manipulations from 1.78 manipulation per object to 0.56 manipulation per object and subsequently reduce the time required from 44.7s per successful manipulation to 17.11s per successful manipulation. However, the single manipulation method experiences drawbacks in terms of its accuracy which suffers from a drop from 97.7% to 77.46% when comparing our proposed single manipulation method to gupta's manipulation aided method.

For future work, we aim fine tune the algorithm to also take into consideration the object of interest location with respect to the work space before deciding a manipulation action to reduce the %error that occurs due to accidental removal of object of interest as well as considering removing the object from scene rather than scene from object when the situation allows it to further improve the efficiency and accuracy of the object isolation not to mention implementing it in a real environment.

ACKNOWLEDGEMENT

Authors are grateful to Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia Melaka (UTeM) for the financial support through RAGS Research Grant funding RAGS/1/2015/TK0/FTK/03/B00118 project. This project was conducted in Center of Excellence in Robotics and Industrial Automation (CERIA) in Universiti Teknikal Malaysia Melaka

REFERENCES

[1] H. Zhang, B. Zhuang and Y. Liu, "Object Classification Based on 3D Point Clouds Covariance Descriptor," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, 2017, pp. 234-237.

[2] Y. Lin; H. Min; H. Zhou; F. Pei, "A Human-Robot-Environent Interactive Reasoning Mechanism for Object Sorting Robot," in IEEE Transactions on Cognitive and Developmental Systems, vol.PP, no.99, pp.1-1

[3] S. Contreras and F. De La Rosa, "Using Deep Learning for Exploration and Recognition of Objects Based on Images," *2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR)*, Recife, 2016, pp. 1-6

[4] S. Elliott, M. Valente and M. Cakmak, "Making objects graspable in confined environments through push and pull manipulation with a tool," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 2016, pp. 4851-4858.

[5] M. Gupta, J. Müller and G. S. Sukhatme, "Using Manipulation Primitives for Object Sorting in Cluttered Environments," in *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 608-614, April 2015.

[6] M. Gupta and G. S. Sukhatme, "Using manipulation primitives for brick sorting in clutter," 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, 2012, pp. 3883-3889.

[7] D. Katz, M. Kazemi, J. A. Bagnell and A. Stentz, "Clearing a pile of unknown objects using interactive perception," 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, 2013, pp. 154-161.

[8] L. Chang, J. R. Smith and D. Fox, "Interactive singulation of objects from a pile," 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, 2012, pp. 3875-3882.

[9] N. M. Ali, S. W. Jun, M. S. Karis, M. M. Ghazaly and M. S. M. Aras, "Object classification and recognition using Bag-of-Words (BoW) model," 2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA), Malacca City, 2016, pp. 216-220.

[10] M. F. Miskon and R. A. Russell, "Mapping normal sensor measurement using regions," 2009 IEEE International Conference on Industrial Technology, Gippsland, VIC, 2009, pp. 1-6.

[11] M. H. Jamaluddin, T. Shimono and N. Motoi, "Motion navigation in haptic bilateral system based on vision-based force compliance controller taking object coordinate into account," 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), Istanbul, 2014, pp. 2238-2243.

[12] M. H. Jamaluddin, T. Shimono and N. Motoi, "An integration method between vision-based disturbance observer and bilateral haptic system for robust tracking of target object," 2014 IEEE 13th International Workshop on Advanced Motion Control (AMC), Yokohama, 2014, pp. 723-728.

[13] M. N. A. H. Sha'abani and M. F. Miskon, "The effect of anomaly detection accuracy in varying the angular resolution of sonar using Repetitive Observation Strategy," 2011 IEEE Symposium on Industrial Electronics and Applications, Langkawi, 2011, pp. 669-674.

[14] C. Dornhege and A. Kleiner, "A frontier-void-based approach for autonomous exploration in 3d," 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics, Kyoto, 2011, pp. 351-356.

[15] T. Maneewarn and P. Rittipravat, "Sorting objects by multiple robots using Voronoi separation and fuzzy control," Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), 2003, pp. 2043-2048 vol.2.

