Methods of Evaluating 3D Perception Systems for Unstructured Autonomous Logistics

Dylan Do Couto — PhD Student
Dr. Joseph Butterfield — First Supervisor
Dr. Adrian Murphy — Second Supervisor
Dr. Joseph Coleman — Corporate Supervisor
Email: {ddocouto01, j.butterfield, a.murphy}@qub.ac.uk, joe.coleman@combilift.com

Abstract

This study introduces methods of evaluating 3D perception systems, such as Time of Flight (ToF) systems, for automated logistics applications in uncontrolled environments. Here perception is defined as a system's understanding of its environment and the Objects Of Interest (OOI) within that environment, through hardware consisting of cameras or depth sensors. Current computer guided machinery that rely on perception systems, such as certain Autonomous Guided Vehicle (AGV), require controlled environments that are specifically designed for such a machine. Uncontrolled environments include warehouses or manufacturing facilities that have not been tailor designed or structured specifically for the purpose of using a computer guided machine. In this study, two methods are proposed to assess 3D systems proposed for autonomous logistics in uncontrolled environments. The results of this study indicate that the methods presented here are suitable for future and comparative 3D perception and evaluation in this space.

1 Introduction

In recent years, the focus on manufacturing and logistics has been the automation of the machinery responsible for the transportation and handling of materials. In autonomous logistics and material handling, there has been a continual improvement of item tracking and computer based logistics [1]. However, this has resulted in the physical machinery and their operation becoming a bottleneck in autonomous systems. The benefits of automating the various material handling machines in these industries are numerous and thoroughly explored, such as improved speed and efficiency of logistics [2], however the requirements for automation in this space are significant. Current State Of the Art (SOA) technology, machines such as forklifts and robotic arms, can be modified to run autonomously outside of structured environments using perception system, but only if key requirements are met. These requirements include the use of "standard loads" consisting OOI that are simple and consistent shapes, such as uniform boxes on pallets. In addition to this, the machines typically operate in isolated areas of warehouses and factories known as restricted zones to prevent both human error from affecting the automated systems and to prevent any injuries from these machines not detecting personnel along the machine path. In this study, a 3D capture system known as a Time of Flight (ToF) camera is examined in the setting of automated logistics. This specific type of 3D capture hardware operates like a conventional camera, using natural light reflected back to a visual sensor that records an RGB colour map known as classical 2D images. For this system however, instead of using natural light the ToF camera projects its own unstructured InfraRed (IR) light source at high speeds intervals, this is used to measure the delay between the source light and the light being reflected in the scene before returning to the visual sensor [3]. This delay determines an object in the scenes distance from the camera, hence the name "Time of Flight". In this study a ToF Basler Blaze intended for autonomous load assessment is utilised, this system presents the recording of high density depth pixels with each capture at high speeds. In comparison to other 3D technologies, such as a Zivid One Structured Light (SL) system, this 3D camera captures footage at a lower dimensional tolerance but a much higher speed (Table 1). In this case, a higher capture speed is more desirable than dimensional accuracy to allow for real time autonomous operations, comparative to a manual machine operator [4]. In the context of this study, noise is presented as white noise or false positives in captured scenes. This may be caused by dust or air particulates as a result of manufacturing processes, such as the use of abrasives, or false positives caused by artifacts of scene illumination as explained further in the methodology.

Table 1: 3D system comparison.

3D system	Spatial Resolution	Temporal noise	Min acquisition time
ToF (Basler)	640x480 points	2 mm	33 ms
SL (Zivid)	1860x1180 points	0.3mm	80 ms

Note: all values presented represent worst case scenarios, i.e. maximum usable distance under worst conditions.

The aim of the work is to establish methods of evaluating 3D systems, particularly the level of white noise present in areas critical to 3D processes, such as edge detection, and the presence of surface distortions, such as depth variations in flat surface or puckering. These are qualities that are vital to the evolution of 3D perception and in this study, methods are presented in both cases for evaluating the presence of these distortions and the degree to which they affect the data. The methods presented here are exclusively tested on one device, a ToF system as described. This is due to limitation with time and equipment available at the time of this study, however the aim of this research is to propose methods of evaluation that are reproducible and can be carried out with various systems in the future.

2 Controlled Environments

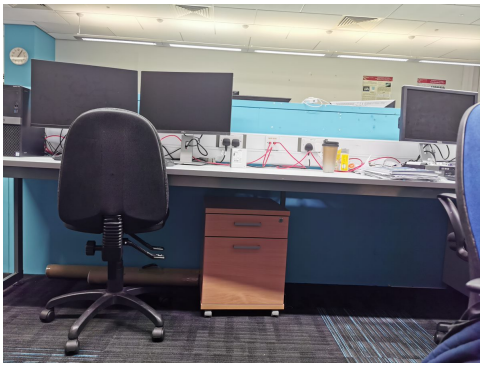

Fig. 1: Office environment QUB

A Controlled environment is an environment where several factors relevant to image processing techniques, classical monocular and/or 3D perceptive methods, are controllable such as lighting, air quality and background objects. This can be as detailed as an environment structured specifically for a given process, such as autonomous bin picking, or as general as a lab setting with consistent lighting. The latter is the case for this study as shown in Fig. 1, where aspects of the environment are consistent allowing for a system to be easily tuned relative to a scenes lighting and objects present in both the foreground and background.

3 Uncontrolled Environments

In contrast to a controlled environment, an un-controlled environment naturally is an environment where variables such as lighting , air pollutants and objects in the scene are neither controllable or consistent. This presents a greater challenge for image processing operations as a perceptive system cannot be finely tuned. For this study a small storage warehouse servicing a forklift factory was

Fig. 2: Warehouse uncontrolled environment

used for capturing data Fig. 2. In this warehouse there was natural lighting provided from sky lights, in addition to this the storage room was located near welding bays which resulted in dense particles circulating in the air as well as numerous unorganised objects present in the background of scenes. This is the primary data set that is used for evaluating the noise the sensor presented in this paper.

4 Methodology

Two main data sets are presented for this study, footage captured in controlled environment represented by the office environment as seen in Fig. 1 and an uncontrolled environment represented by footage captured in an uncontrolled warehouse environment Fig. 2. From each data set, two aspects of the objects present in the scene will be evaluated; the noise present along an objects edge, in particular with gray scale depth images captured in the uncontrolled environment, and surface distortions observed in point cloud data, examined through flat surface both in controlled and uncontrolled environments. To examine these aspects of the data, two methods are presented below.

4.1 Extraction by Threshold

Gray scale thresholds are a common tool used in classical image processing of 2D images. As presented in early literature, thresholds are typically used for segmentation in 2D gray scale images as described by Cheriet et al [5]. The process operates by determining an images background and foreground objects based on their colour, or gray scale value between 0 & 255 (256 bit resolution). This can be as simple as applying a simple cut-off value, where the gray scale values beyond a stated value are ignored ,i.e. pixels with a value higher than 125. More complicated methods exist such as Otsu threshold criteria, however these are adaptive thresholds that are not applicable in this study. Consider the image below of a standard rectangular load captured through 2D IR data returned, a typical 2D image obtained through IR light instead of natural light.

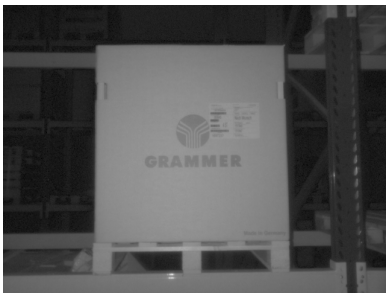

Fig. 3: Gray scale image of OOI

Observing Fig. 3 it is reasonably easy to distinguish the background in the image, as it is much darker than the pallet load and the racking. However, the gray scale difference between racking present in the foreground and background is a harder distinction to be made. Depending on the purpose of applying a threshold filter to an image, the tuning of the threshold might need to be broad and adaptive, as discussed with Otsu tuning, which accounts for a scene where the objects present or the lighting of the scene will vary. In this case, the threshold is manually tuned to extract only the OOI, the pallet load, as small data set is being observed. Applying this simple binary threshold yields the following result Fig. 3.

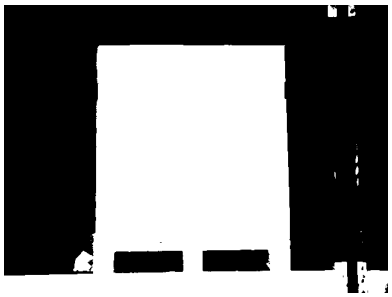

Fig. 4: Threshold image of OOI

The result of this threshold operation yields a binary image of only the OOI, however there are other artifacts present in the image from the objects racking some background surfaces. This results in a base image that is not ideal, however the purpose of this image is to complete a comparison of data present in the gray scale depth map which allows for minor defects to be present. This is expanded upon later in the chapter.

Fig. 5: Threshold image of OOI

In addition to the 2D IR image obtained, a gray scale depth map was also captured. This is generated by measuring the time delay between a pulse of IR light and the light returning to the 3D camera, as discussed in the introduction. This is represented as a gray scale image, where each pixel value between 0 and 255 represents a detected points distance from the 3D camera and not its colour. As a result, a similar threshold technique can be applied to depth images to extract data that is a specific distance from the camera. This is seen in Fig. 5, where OOI has been extracted from the background and isolated. As a result the OOI is obtained two ways, an IR threshold that extracts only the OOI with minimal noise and the Depth map threshold that presents the OOI and only the relevant noise along the edge. To examine the degree of noise present in the depth map, a direct comparison can take place by performing a bitwise subtraction of the base image, IR threshold, and the depth map leaving only the noise. This operation is completed by only removing pixels that are present in both image, where noise is only present in the depth map, and defects in the IR threshold are ignored as this is the base image. Only pixels in the depth map are carried through after a bitwise subtraction unless they are also present in the base IR image. This is the method proposed for detecting noise edge for OOI presented in uncontrollable scenes.

4.2 Plane normal vectors

Surface properties, specifically surface normal's, are a key property of point clouds that are used for OOI separation in complex sciences, as presented in Tsai et al [6]. There are varies methods and algorithms that are used to determine plane normal vectors from a point cloud surface, in this study the "Open3D" libraries were used. This algorithm considers a close neighbourhood of points, points within a stated radius, and results in a set of normal vectors that are calculated incrementally along a surface and represent the direction the surface is facing at each point, Fig. 6.

For an ideal flat surface, the normal vector is represented as [0,0,-1]. This indicates a vector that is pointing directly at origin, i.e.

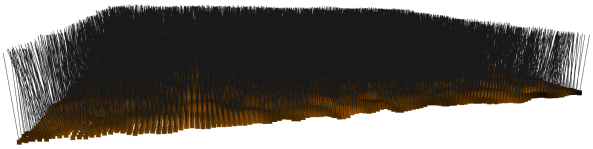

Fig. 6: point normal vector map

the 3D camera. Points that indicated a puckered or warped surface would present a small deviation from this value, e.g. [0.210,-0.017,-0.978]. By calculating the distance between the end point the ideal normal vector and the actual normal vector Eq 1, a metric can be derived to determine the degree of which the surface is warped.

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2} \quad (1)$$

By comparing this distance for all points in a given area along a surface, a standard distribution can be developed that presents the mean normal distance/surface error that exists for a given surface and the standard distribution of normal vector distance/surface error Fig. 7.

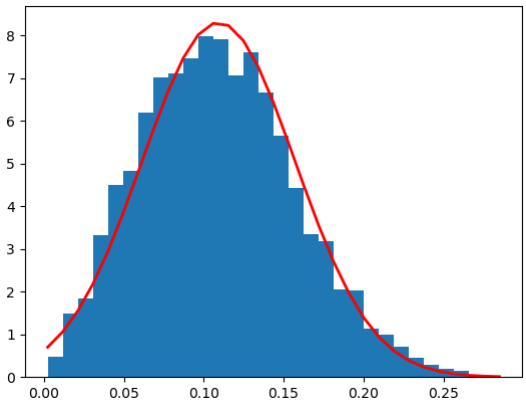

Fig. 7: Distribution of normal vector error

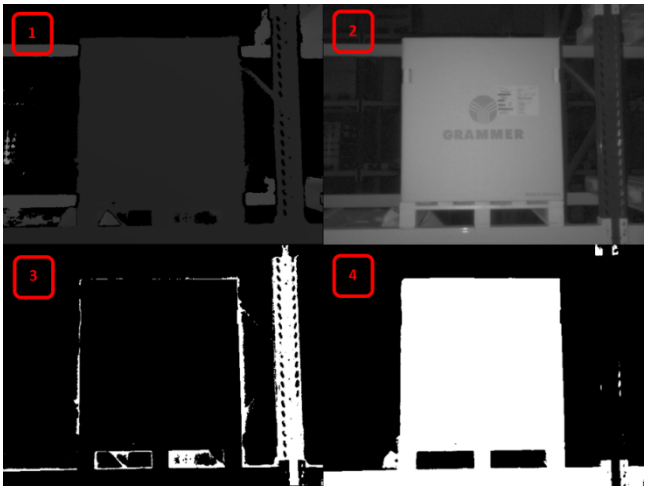

Fig. 8: Processing of noise in depth data

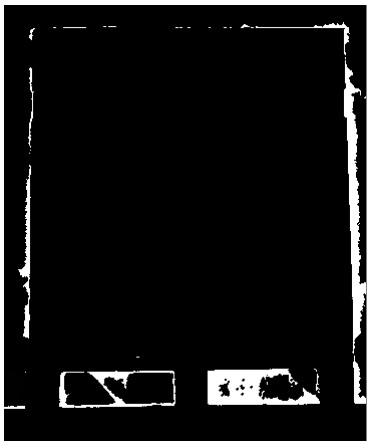

Fig. 9: Isolated noise in depth data

5 Results

5.1 Edge Distortions

As discussed in the method section, noise along the edge of OOI's Fig. 9 can be quantified by comparing the gray-scale depth data with that of an tuned threshold system for extracting the OOI and creating a mask to perform a bit-wise subtraction. As the ToF system can capture both a gray-scale depth Fig. 8.1 and intensity image Fig. 8.2 simultaneously, this allows for a comparison between classical methods and depth data to take place. Through using classical methods, the OOI is extracted by using an optimised gray-scale threshold to isolate the load, pallet and rack based on its gray value Fig. 8.4. By preforming the same type of threshold, the gray-scale depth image can be altered to contain only the OOIs and the noise present. Performing a simple subtraction process by subtracting the white pixels from the binary images of both the depth map and the 2d gray-scale image (BITWISE NOT) leaves only the noise present Fig. 8.3.

However, due to the difference between the depth and intensity image other undesirable objects remain in the frame. To remove these a simple bounding rectangle was drawn around the largest cluster of White pixels in the mask image Fig. 3.4 and used to crop the result Fig. 3.3, which creates an image consisting solely of the noise pixels which can be quantified by counting the number of white pixels present Fig. 4.

However, this value by itself means nothing as the total number of pixels can vary between scenes in addition to objects being closer or further away. As a result, we can determine a value of noise proportional to the OOI in the scene by quantifying it as a ratio of total pixels in the cropped image (157,108 pixels) and the noise present (8566 pixels). This give a noise ratio of 5.45%. Naturally this value does not determine the difficultly in segmenting and analysing the OOI, but it can be used as metric to compare the level of noise present across different data sets. If the same process is completed for the isometric point cloud data, the level of noise found can be used to determine if a reduction in noise can be achieved by simply altering the position of the capturing system. Completing the same process as before the OOI is isolated and a mask is used to subtract items from the scene so that only the noise is present Fig. 10.3.

As a result of the data being captured in an isometric perspective, only 3 of the 4 edges can be examined as the noise from the left edge will blend with the left surface that is visible. This can be seen in Fig. 11 below which had a noise ratio of 5.07% (5301/104520 pixels). Clearly visible in Fig. 11, there is less noise in the image most notably on the edges of the OOI and the racking beam holding the pallet. However, the noise is only reduced by 7% (5.07/5.45 * 100).

This may be attributed to an increase in noise in the inner section of the pallet, or simply that re-positioning the depth camera has only a minor effect on noise. In any case, a reduction in noise can be seen showing some benefit to an indirect positioning of the 3D sensor however limited that reduction may be.

5.2 Surface Distortions

The last method for assessing the sensor presented in this study is to determine the effect of surface defects. When determining surfaces to complete scene segmentation, smooth and consistent point clouds that represent these surfaces are essential. Variance in the depth values of points recorded can be expected, however this variance should be below an acceptable threshold to not affect algorithms that depend on surface properties of point clouds or meshes. To measure the effect of surface distortions, scans were taken of two scenes perpendicular to the OOI's and of flat surfaces to allow for easy assessment of surface distortions. The first scan

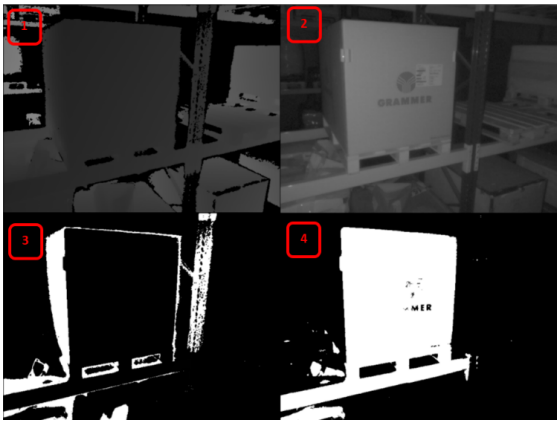

Fig. 10: Isometric data set

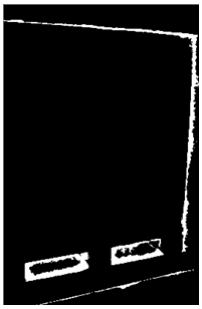

Fig. 11: Noise of isometric data set

was taken of the same un-structured scene presented in Fig. 12, taken in a relatively open space with natural lighting through sky lights. At the time of capturing these data set, the natural lighting was brighter than normal (brighter than 90,000 lux) in comparison to the recommended warehouse illumination from overhead lights (10,000 – 20,000), producing a challenge for the system to Suppress Background Illumination (SBI), this however results in a noisier surface, as can be seen bellow. As seen in the point cloud, the flat surface of the cardboard box is heavily dimpled and there is severe noise long the edge of the load. This noise presents several challenges for processing the load and environment such as edge finding for object template matching and determination of clearances. For comparison a similar capture was taken in an office setting using only indoor lighting at a slightly further distance. If we crop a section of this point cloud to include only the flat surface of the cardboard box, the surface defects can be easily seen.

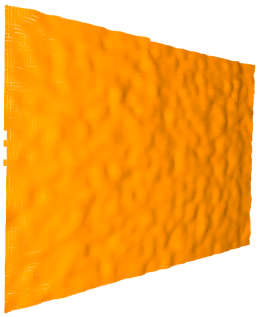

Fig. 12: Cropped point cloud of flat surface (un-structured environment)

Using this cropped section, the normal vector for a given radius of points, in this case a radius of 10 points and a maximum nearest neighbor cluster of 30, can be calculated for each point. This vector gives an indication to where the surface is facing in 3D space relative to the capture source of the point cloud. For the OOI captured, we know the front surface of the OOI is perpendicular to the sensor and so the normal vector value for a flat surface is [0,0,-1]. With this value known, the distance of the returned normal vector for each point and the ideal point, of [0,0,-1], can be measured and collated as a metric to determine useful statistics of the surface captured, such as the mean distance from the target and the standard deviation. Comparing the surface captured in the un-structured environment to the structured office environment, the same type of flat surface perpendicular to the camera can be observed in the drawers under the desk. This OOI is at a comparable distance to the previous OOI and has a large flat surface. Less noise is seen in this seen around the edges of the OOI however, dimpling can still be seen along the flat surface.

Performing the same steps on this data set, the normal of the surface can be measured for a flat section of the lower drawer and the variance in this value can be calculated. Comparing the mean error, the distance in 3D space between the ideal end point of the normal vector and the actual end point, a clear improvement in the captured surfaces compared to the unstructured environment.

Table 2: Recorded surface error.

Scene	Mean Error	Standard Deviation
Warehouse	0.124	0.0058
Office	0.109	0.0048

6 Conclusion

In this study two aspects of 3D data was investigated, noise along the edge of OOIs and surface distortions, particularly puckering of flat surfaces. To measure edge noise, colour based gray scale images were captured simultaneously with depth maps and were compared using classical image processing methods. By applying a threshold to each data type to extract only the OOI and performing a bitwise subtraction of the two results, the noise present along the OOIs edge in the depth map could be isolated. It was found that capturing OOI in an isometric angle opposed to perpendicular to the surface results in a small noise reduction of 7%. Although this result shows a clear improvement, other methods of improving the capture of OOI edges, such as optimizing the capture settings if a consistent environment is used. However, this does confirm that this is a viable method for assessing both the quality of captured data and sensor that is used for capturing scenes in this application.

Surface distortions were evaluated by calculating the surface normal vectors of a large flat surface of an OOI and comparing this to the ideal normal vector (0, 0, 1) to determine the statistical error, the mean and standard deviation from the ideal normal vector. The surface normal is outline in this report as one of the key methods for determine the surfaces that comprise faces of an OOI and a consistent value results in faster plane detection. By completing this analysis on the limited data set provided, a clear improvement was seen between surface captured under controlled lighting, such as office lights, and uncontrolled lighting, such as sunlight, as seen in Table 2. This comparison not only shows the challenges presented of capturing flat surfaces in an uncontrolled environment, but also that this is a viable method of measuring any variance in surface distortions.

References

[1] G. Q. Huang, M. Z. Q. Chen, and J. Pan, “Robotics in ecommerce logistics,” *HKIE Transactions*, vol. 22, no. 2, pp. 68–77, 2015. [Online]. Available: https://doi.org/10.1080/1023697X.2015.1043960

[2] J. Wang, N. Zhang, and Q. He, “Application of automated warehouse logistics in manufacturing industry,” in *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, vol. 4, 2009, pp. 217–220.

[3] R. Lange and P. Seitz, “Solid-state time-of-flight range camera,” *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[4] S. Zhang, “High-speed 3d shape measurement with structured light methods: A review,” *Optics and Lasers in Engineering*,

vol. 106, pp. 119 – 131, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0143816617313246

[5] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 918–921, 1998.

[6] C. Tsai and S. Tsai, "Simultaneous 3d object recognition and pose estimation based on rgb-d images," *IEEE Access*, vol. 6, pp. 28 859–28 869, 2018.