Development of a Single Camera Machine Vision System for Automatic 3D Size Detection

Y. H. Tay,1, U. Khairuddin.1*
1Department of Mechanical Precision Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.

*Corresponding author: uswah.kl@utm.my

Abstract. The rapid growth of global e-commerce prompts the need of efficient warehouse handling and logistics, forcing manual operations to be replaced with automatic systems. An example of automated solutions is the smart packaging system for boxes which can simulate optimized box arrangements in order to save space. Even though three-dimensional bin packing problem has been studied widely to optimize box arrangement, only a few studies have been done to automatically detect box sizes. Box size detection is important to increase the effectiveness and efficiency of the packaging process as well as providing fast and accurate input for the box arrangement optimizer. Therefore, this paper presents the development of a machine vision system for automatic box size detection in 3D to support a smart packing simulator. The system uses a single camera and a platform covered with square grids prints. The box size detection algorithm was based on the localization of the platform area and it works by applying the bilateral filtering, binary thresholding and morphological image transform for the square grids feature extraction. To measure the box size, the length, width, and height of the boxes were detected by referencing it to the count of the square grids. The volume of the boxes was further computed from the three-dimensional values obtained. The performance of the algorithm was then evaluated by calculating the error against the true value obtained from the manual measurements. The average accuracy for box size detection was 94.3% and analysis shows that the accuracy of the model was highly dependent on the size of the square grids on the platform. The result shows great potential of using a single camera system for automated 3D box size detection.

1. Introduction
E-commerce online shopping is having a significant impact on the logistics process or more specifically transportation, warehousing, and packaging [1]. Among the automated warehousing equipment, automatic box size detection is important as it can reduce labour costs for manual handling activities, reduce the time needed for packing the purchased items and minimize solid waste generated from the packaging process [2]. Machine vision system has been widely used in the manufacturing sector for tracing the manufactured products, and in the agriculture sector for robot harvesting. However, when it comes to the packaging sector, the application of the machine vision system is still very limited even though there is a high demand for the technology. Therefore, it is important to develop a machine vision system for automatic box size detection.
Generally, researchers have localized the targeted object area and eliminated the background area for the size detection. Most researchers [3-10] perform image segmentation to separate the targeted object from the background by applying the image processing algorithm. Some image processing techniques used for image segmentation are Weiner Filtering [3], Gaussian Kernel Filter [6], Otsu Threshholding [3,4], Harris Corner Detection [7] and Circular Hough Transform [3]. However, the outcome of image segmentation is heavily dependent on the surrounding lighting condition. On the other hand, recent studies [5] utilize multiple hardware or devices to obtain multi-perspective images. The system composition includes multiple cameras in different axes and a turntable [5]. Most of the studies can only detect the object size in three dimensions by using multi-perspective images which will be resulting in high tooling cost.

This paper presents the development of a single camera machine vision system for automatic 3D box size detection. The machine vision system consists of a camera and a platform covered with square grids prints. The box size detection algorithm will be based on the localization of the platform area to detect box size by referencing it to the square grids counts.

2. Methodology
The box size detection algorithm is designed to localize the platform area by applying the image processing techniques for the square grids feature extraction.

2.1. Box size detection algorithm
The datasets used in the system are the square grids attached on the platform that act as a background in the image. The methodology can be divided into 3 Phases. Phase 1 which is the preparation phase includes the setup of the image acquisition system. In Phase 2, image processing is done to enhance the image quality and followed by image segmentation to eliminate the box area. After that, feature extraction is done to detect the square grids in the platform area. The number of square grids detected is correlated into length, width and height of the box. Phase 3 covers the evaluation on the size detection algorithm against the manual measurement method to obtain the accuracy of the system. The workflow of the simulator is in Figure 1.

![Flowchart of the box size detection algorithm](image-url)
2.1.1. **Image acquisition system.**
The image acquisition system captures the images of the box and the platform covered by printed square grids and transfers it into a computer. The constraints for this setup includes the external lighting environment must be controlled in a colour temperature of about 7300K, the camera must be placed at 5cm to 10cm from the platform, and the box must be placed at the leftmost side of the square grids area.

2.1.2. **Image processing.**
The box size detection algorithm converts the colour image into a grayscale image in order to reduce the complexity of image processing and to reduce the processing time. Then, it applies a bilateral filter to remove the noise and smooth the image while keeping the edges sharp. Binarization of the image is performed to identify the Region of Interest (ROI) by selecting the pixel of interest according to the light intensity. The algorithm applies binary thresholding to convert the grayscale image into a binary image at a threshold value of 170. After that, it implements morphological erosion to thicken the border of the square grids for improving the square grids feature extraction. Figure 2 shows the transformation of a colour image into a binary image after performing morphological erosion.

![Figure 2](image-url)

**Figure 2.** Image transformation; (a) RGB image, (b) Grayscale image, (c) Binary image, and (d) Binary image after performing morphological erosion

2.1.3. **Feature extraction.**
The algorithm utilizes the contour feature to detect the square grids on each platform. By setting a range of contour area, the algorithm is able to detect the square grids when the area bounded by the border fall within the limits. The minimum and maximum contour area for the vertical platform is set to be 625 and 1800 respectively, while for the horizontal platform it was set to be 400 and 2200
respectively. Figure 3 shows the detected square grids on the platform. The algorithm then correlates the square grids counts to the box dimensions by the mathematical equation.

![Figure 3. Square grids detection on the platform](image)

2.1.4. Evaluation.
The detected dimension is compared to the measured dimension to obtain the error by using Mean Absolute Percentage Error, MAPE as shown as follows:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \frac{|L_a - L_d|}{L_a}$$

where

- $L_a$ = the actual dimension of the box
- $L_d$ = the detected dimension of the box
- $N$ = total number of box samples

3. Result and discussion
The proposed algorithm was tested on thirty boxes of different size. The actual sizes of the boxes were manually measured using a ruler. The performance of the system was obtained by comparing the system’s output and the actual box size. Table 1 shows the percentage error for each dimension and the volume of the box with minimum error of 0% and maximum error of 139.66 % (this measure for 81mm). The overall accuracy achieved was 85.19%.

| Box | L (%) | W (%) | H (%) | Vol. (%) | Box | L (%) | W (%) | H (%) | Vol. (%) |
|-----|-------|-------|-------|---------|-----|-------|-------|-------|---------|
| 1   | 0.00  | 5.56  | 1.35  | 6.02    | 16  | 18.87 | 15.00 | 20.97 | 11.01   |
| 2   | 10.53 | 4.30  | 3.25  | 12.41   | 17  | 7.27  | 0.00  | 0.58  | 8.10    |
| 3   | 55.17 | 13.60 | 139.66| 146.77  | 18  | 4.76  | 2.17  | 3.56  | 7.28    |
| 4   | 5.56  | 0.00  | 1.11  | 6.71    | 19  | 4.08  | 21.11 | 22.16 | 36.04   |
| 5   | 7.50  | 6.67  | 6.92  | 7.76    | 20  | 17.74 | 35.79 | 7.82  | 20.15   |
| 6   | 9.09  | 3.03  | 0.86  | 6.66    | 21  | 7.69  | 1.94  | 0.55  | 10.85   |
| 7   | 5.56  | 2.94  | 3.41  | 5.92    | 22  | 5.56  | 20.59 | 47.02 | 67.44   |
| 8   | 0.00  | 2.86  | 12.14 | 14.48   | 23  | 37.04 | 49.52 | 43.61 | 35.59   |
| 9   | 2.86  | 0.00  | 11.48 | 13.77   | 24  | 21.21 | 10.91 | 5.58  | 9.77    |
| 10  | 4.23  | 2.70  | 3.50  | 4.27    | 25  | 1.92  | 11.61 | 16.13 | 26.98   |
| 11  | 13.04 | 1.30  | 1.16  | 8.87    | 26  | 1.64  | 6.52  | 2.33  | 1.18    |
| 12  | 6.85  | 1.33  | 6.13  | 13.13   | 27  | 4.44  | 0.00  | 19.71 | 24.50   |
Generally, the box size detected using the algorithm was underestimated when compared to the actual box size. The major sources of error included inaccurate square grids detection and error in the box segmentation. Error in square grids detection occurred when the system failed to extract the square grids feature. It often appeared in white coloured box where the edge of the box appeared to fuse with the square grids as shown in Figure 4. The border of the square grids was disappeared, causing the algorithm failed to detect the square grids using the contour feature. At the same time, the symbol and texture on the box surface might be detected as a unit of square grid if the contour area were coincidently fell within the range of contour area defined in the algorithm.

The proposed algorithm was then tested for thirty boxes and the dimensions are the same as the previous case, but each box is wrapped with a non-reflective black coloured paper. Table 2 shows the percentage error for each dimension and the volume of the box with minimum error of 0% and maximum error of 44.71% (this measure for 38mm). The overall accuracy achieved was 94.35%.

| Box | L (%) | W (%) | H (%) | Vol. (%) | Box | L (%) | W (%) | H (%) | Vol. (%) |
|-----|-------|-------|-------|----------|-----|-------|-------|-------|----------|
| 1   | 0.00  | 5.56  | 3.38  | 4.24     | 16  | 3.77  | 5.00  | 2.32  | 3.39     |
| 2   | 10.53 | 2.38  | 15.45 | 4.80     | 17  | 7.27  | 4.91  | 4.09  | 1.57     |
| 3   | 12.07 | 5.17  | 3.45  | 5.26     | 18  | 4.76  | 3.26  | 1.78  | 4.50     |
| 4   | 5.56  | 6.25  | 4.44  | 4.25     | 19  | 12.24 | 0.00  | 15.91 | 0.65     |
| 5   | 15.00 | 3.33  | 1.54  | 14.38    | 20  | 3.23  | 8.42  | 4.47  | 8.93     |
| 6   | 9.85  | 4.55  | 2.59  | 3.01     | 21  | 7.69  | 0.00  | 2.19  | 6.27     |
| 7   | 5.56  | 0.00  | 1.14  | 4.21     | 22  | 5.56  | 0.00  | 1.19  | 4.43     |
| 8   | 11.76 | 0.00  | 2.86  | 14.48    | 23  | 3.70  | 0.00  | 2.26  | 3.58     |
| 9   | 14.29 | 4.00  | 16.39 | 2.27     | 24  | 3.03  | 1.82  | 0.93  | 0.33     |
| 10  | 4.23  | 5.41  | 1.50  | 10.80    | 25  | 1.92  | 1.79  | 2.58  | 0.95     |
| 11  | 13.04 | 1.30  | 1.16  | 8.87     | 26  | 1.64  | 0.72  | 0.58  | 3.46     |
| 12  | 6.85  | 1.33  | 0.77  | 4.08     | 27  | 4.44  | 1.29  | 1.92  | 8.57     |
| 13  | 10.53 | 44.71 | 30.70 | 69.28    | 28  | 1.92  | 2.50  | 0.90  | 5.33     |
| 14  | 3.38  | 1.04  | 4.94  | 2.80     | 29  | 5.11  | 2.89  | 2.31  | 2.99     |
| 15  | 2.38  | 3.30  | 1.59  | 0.53     | 30  | 5.56  | 1.11  | 0.47  | 6.12     |

Mean L (%) = 6.77 W (%) =4.07 H (%) = 4.56 Vol (%) = 7.21

Table 1. Percentage error for each dimension and the volume of the box

Table 2. Percentage error for each dimension and the volume of the box

The box size detection accuracy based on the boxes wrapped by black coloured papers was substantially higher in comparison to the boxes without wrapping. After wrapping the boxes with a
non-reflective black coloured paper, the borders of the square grids were easily detected. The absence of texture on the box surface helped to prevent the miscount on number of square grids. The improvement on square grids detection and box segmentation as shown in Figure 5 was the key for the increment in the box size detection accuracy. In this case, the accuracy of the box size detection algorithm was limited by the size of the square grids which represent the sensitivity of the system.

![Figure 5. The detected square grids (left) and error in square grids detection (right)](image)

4. Conclusion
This paper proposed a single camera machine vision system for automatic 3D box size detection. The system computes the box size from a single perspective image by detecting the square grids count on the platform. The automated process made the system very fast and has 94.35% accuracy and may be the potential solution to the packaging industry.

In future work, more containers shape such as cylinder and sphere should be added to provide a better support for the packaging process. Also, the accuracy of the system can be improved by reducing the square grids size and increasing the resolution of the camera.

References
[1] Delfmann, W., Albers, S., & Gehring, M. (2002). The impact of electronic commerce on logistics service providers. International Journal of Physical Distribution and Logistics Management, 32(3), 203–222
[2] Hellström, D., & Saghir, M. (2007). Packaging and logistics interactions in retail supply chains. Packaging Technology and Science, 20(3), 197–216
[3] Gongal, A., Karkee, M., & Amatya, S. (2018). Apple fruit size estimation using a 3D machine vision system. Information Processing in Agriculture, 5(4), 498–503
[4] Nyalala, I., Okinda, C., Nyalala, L., Makange, N., Chao, Q., Chao, L., … Chen, K. (2019). Tomato volume and mass estimation using computer vision and machine learning algorithms: Cherry tomato model. Journal of Food Engineering, 263(July), 288–298
[5] B. Dellen and I.A. Rojas. (2013). Volume measurement with a consumer depth camera based on structured infrared light. 16th Catalan Conference on Artificial Intelligence, m, 1–10
[6] Eifert, J. (2003). Fast surface approximation for volume and surface area measurements using distance transform. Optical Engineering, 42(10), 294
[7] Quintino Ferreira, B., Griné, M., Gameiro, D., Costeira, J. P., & Sousa Santos, B. (2014). VOLUMNECT: measuring volumes with Kinect. Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2014, 9013, 901304
[8] Xu, Y., & Brownjohn, J. M. W. (2018). Review of machine-vision based methodologies for displacement measurement in civil structures. Journal of Civil Structural Health Monitoring, 8(1), 91–110
[9] White, D. J., Svellingen, C., & Strachan, N. J. C. (2006). Automated measurement of species
and length of fish by computer vision. Fisheries Research, 80(2–3), 203–210

[10] Fernandes, L. A. F., Oliveira, M. M., da Silva, R., & Crespo, G. J. (2006). A fast and accurate approach for computing the dimensions of boxes from single perspective images. Journal of the Brazilian Computer Society, 12(2), 19–30