4th International Congress of Developing Agriculture,
Natural Resources, Environment and Tourism of Iran

13-15 Feb. 2019, Tabriz Islamic Art University
In cooperation with Shiraz University and Yasouj University

شماره مجوز کنگره از وزارت علوم: ۷۶۸۶۸۶۹

شماره مجـوز نمایه در پایگاه استنادی علوم جهان اسلام ۷۰۱۰۱ - ۹۷۱۸۰

On the Feasibility of Estimating Fruits Weights Using Depth Sensors

Seyed Muhammad Hossein Mousavi*, V. B. Surya Prasath

1. Department of Computer Engineering, Faculty of Engineering, Bu Ali Sina University, Hamadan, Iran, corresponding author : mosavi.a.i.buali@gmail.com*
2. Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati OH 45229 USA, surya.prasath@cchmc.org

Abstract

Automatic fruit weight estimation is an essential in today's automated fruit and agriculture industry. To increase the weighing efficiency of fruits, and to decrease the usage various tools that require dedicated human efforts, robust and easy to use automatic machine vision-based systems are required. Such systems can further reduce human errors, willful manipulation at point of sale counters in traditional shops in weighing fruits as well as decrease the overall costs of manual systems. Despite a wealth works in automatic fruit inspection systems with image processing techniques, there are not many works that adopt consumer depth cameras like Microsoft Kinect. In this work, we study the feasibility of an automatic fruit weight estimation method for Sweet Lemons (Citrus limetta), Sweet Peppers (Capsicum annuum), and Tomatoes (Solanum lycopersicum) based on their color (RGB) and depth images. Given the lack of available public datasets for this research direction, we create a dedicated database that consist of color plus depth information using data from Microsoft Kinect V.2 sensors with 50 samples in each of the three fruits. Our novel method is evaluated using the quality metrics such as mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE), which are calculated on three different distances, namely 0.8, 1.0, and 1.3 meters, from the sensor. One of the main goals of this paper is to study whether such depth sensor-based weighing method can remove the need for cumbersome industrial machines. Our proposed method is a non-learning-based system; hence it is fast and potentially could be used on real time automatic inspection systems. Our feasibility study indicates we obtain satisfactory accuracy in weighing these fruits and provides a potential for other fruits.

Key words: Automatic fruit weight estimation, Image processing, Depth data, RGB-D, Automatic inspection, Low cost, Microsoft Kinect

1. Introduction

Automatic image processing techniques are gaining popularity in the agriculture industry. Automatic fruits weighting and grading is part of industry 4.0 and there is various specialized equipment available in the commercial market. Typical automatic fruit grading machines are specially designed to sort and grade particular kinds of fruits and vegetables according to the weight. They are widely used for grading apples, pears, onions, lemons, potatoes and other round fruits and vegetables [1]. Automatic image processing-based systems can weigh the fruits in an automatic fashion [2]. These automatic expert systems can be used to assist or replace human efforts, due to they are much more efficient, faster, and precise with less or no errors [3]. Figure 1 shows some of the available industrial fruit weighting and grading machines.

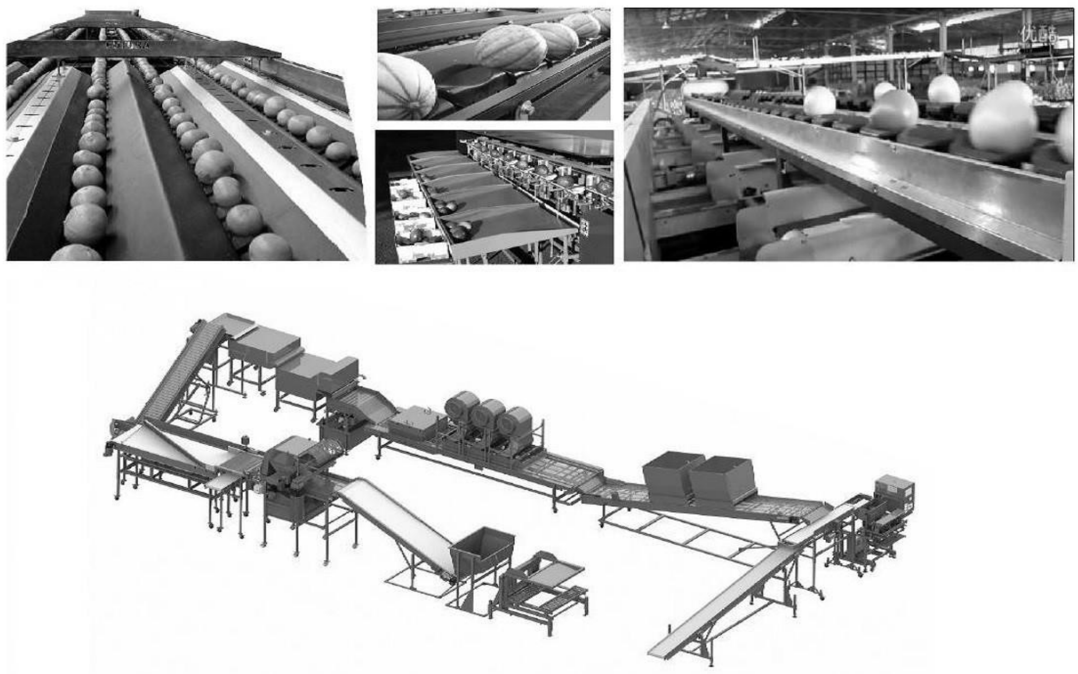

Figure 1 : Industrial fruit weighting and grading machines. Adding cost effective automatic depth sensors in the grading system can alleviate the burden of manual labor.

The available image processing techniques for fruit weight estimation use mainly the intensity or color (RGB) values to identify and estimate weights. Recently, the availability of commercial depth sensors, like Microsoft Kinect, that are cheap and easy to use have opened the way to utilize the depth information along with intensity/color pixel values for automatic image processing methods. A RGB-D image is simply a combination of a RGB image and its corresponding depth image. A depth image is an image channel in which each pixel relates to a distance between the image plane and the corresponding object in the RGB image. It is also termed as 2.5D or Range image [4]. In 3D computer graphics, a depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint. The term is related to and may be analogous to depth buffer, Z-buffer, Z-buffering and Z-depth [5]. Having depth image, it is possible to make 3D form of the object. Using depth sensors for calculating the distance between robot and objects in robotic is very common these days [6]. Figure 2 presents one of the samples of vegetables, an eggplant, in color and depth form, where we have highlighted three different values in the depth image representation to highlight how it captures the distance from sensor. As can be seen, the infrared spectrum in the depth sensor data reaches to the sides of the eggplant, the values get bigger indicating the depth edges are farther, and at the middle of the vegetable the values are smaller and is closer to the sensor.

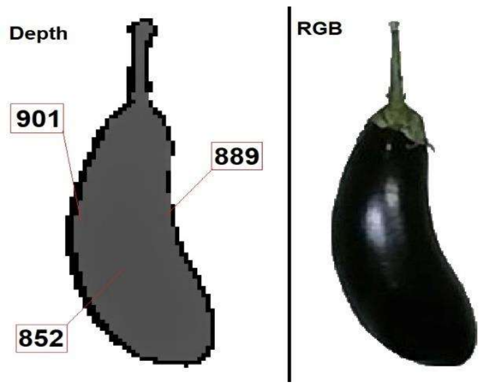

Figure 2 : Depth sensors obtain color + depth (RGB-D) data. Left image is the depth structure shown as an intensity channel of an eggplant vegetable. The smaller value means closer distance to the sensor and vice versa. Right image is the standard color (RGB) image of the same.

Among the available commercial depth sensors, Microsoft Kinect is one of the most used depth sensors to date. It can be used on Microsoft Xbox 360 (Kinect V.1) or Xbox one (Kinect V.2) consoles or be used as a developer device. Kinect 2.0 was released with Xbox One on November 22, 2013, see Figure 3. Because of the lower price and high efficiency, a lot of developers and researchers have explored of their applications in various areas. The sensor can record RGB and depth video frames with 1920*1080 resolution for RGB images and 512*424 for depth images at 30 fps. Further, it is capable to work between 0.8 to 5.0 meters ranges [7]. Kinect sensor could be used in different fields of industry [8], robotics [9], 3D games and animation [10], in psychology like facial expression recognition [11], gesture recognition [12], action recognition [13], and security [14]. The advantage of such a depth sensor is in providing extra information about the object that is being imaged, which is another dimension (depth) that leads to higher accuracy and less error in further object identification and analysis.
Research on automatically estimating fruits weights using depth sensors is in a nascent, studying a system for this purpose is essential. In general, using only the color information to estimating fruits weights lead to high percentage of estimation errors at different distances. These errors happen primarily due to a lack of distance understanding between the imaging sensors and the fruits. For example, let us say we are calculating automatically the weight of an apple with color images in two different distances of 1 and 2 meters. Due to the imaged apple's size changes, it is necessary to calculate each of these sensor-object distances precisely. Further, if the distance is changed the color image processing techniques need to be adjusted for the distance changes. However, if the distance data between object and sensor exists, i.e. the depth

data, calculating the distance between object and sensor could be obtained automatically, and objects weight can be calculated efficiently. Major advantages of using depth sensors includes the avoidance of sophisticated industrial robots for calculating the weight in a normalized distance, and the possibility of discerning the fruits weights in flexible distances, for example 0.8 to 5.0 meters for Kinect V.2 depth sensor. In this work, we utilize the depth sensor to estimate the fruits weights based on preprocessing, pre-segmentation, and distance from the sensor. Our proposed approach, an automatic fruits weight estimation technique, explores the feasibility of using image processing techniques for estimating the weights quickly with depth sensors. Our experimental results on a set of sweet lemons, sweet peppers, and tomatoes with Kinect V.2 depth sensor indicates that we can obtain reliable weight estimations by incorporating color and depth information.
Rest of this work is organized as follows. In section 2, related literature in automatic fruit weight estimation is discussed briefly, and the proposed method using color and depth images is presented in detail. Section 3 includes proposed fruit database and evaluation results with various error metrics at different distances (object from sensor). Finally, Section 4 concludes the paper.

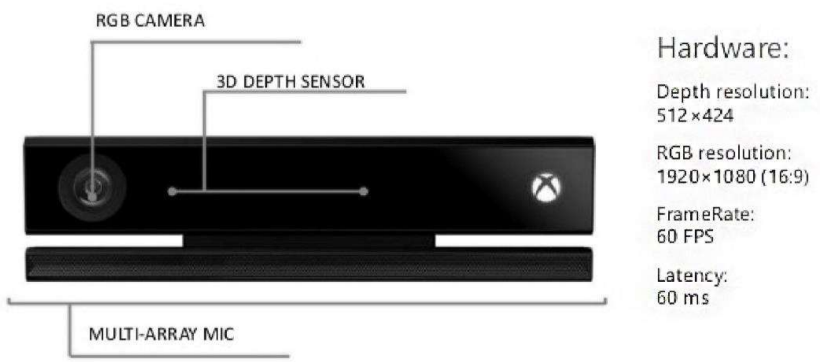

Figure 3 : Kinect V.2 depth sensor and its specifications.

2. Automatic image analysis methods for fruits weights estimation

2-1- Previous works

Teoh and Syaifudin [15] used image processing and analysis techniques for estimating weight of Chokanan mangoes. They used 50 samples and linear regressions for evaluation and reached 0.9769 correlation coefficient. Also mean absolute error (MAE) was 3.76 for all the samples. They used pixel counting for estimating the weight.
Nuske [16] made a system for yield estimation in vineyards by visual grape detection. They counted the grape clusters and calculated its corresponding weight. They employed 244 samples for two different grape type and reached 9.8% of mean error.
Font et al. [17] proposed a system for vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. They employed RGB and HSV color channels and reached 16% and 17% error for each channel respectively. It should be noted

that they used a learning-based approach for grape weight estimation. They used a pixel-based segmentation step along with Bayesian classifier for the classification task.

Omid et al [18] used image processing techniques to make a system for estimating volume and mass of citrus fruits. They used lemons, limes, oranges, and tangerines as samples. The coefficient of determination (R^2) for lemon, lime, orange, and tangerine were 0.962, 0.970, 0.985, and 0.959, respectively. They showed that the volume and mass are highly correlated in various citrus fruits. They captured data in two different view and measuring was based on digital balance method. Further, they used a lot of image preprocessing techniques like background segmentation, de-noising etc. Finally, they employed linear regression to evaluate results. They used visual basic programming language for implementation.

Spreer and Müller [19] made a system to estimating the mass of mango fruit from its geometric dimensions by optical measurement. They used three geometric dimensions to calculating mass. They used 30 mango for validation part and reached root mean standard error of 12.22. Further, a high correlation ($R^2 = 0.96$) was found between measured and calculated mass.

Marinello et al [20] used Kinect V.1 to estimate grape mass. They reconstruct 3D shape of the grapes and tried to calculate the mass through 3D model. They utilized Kinect sensor between 0.5-1.0 ranges and reached well ($R^2 = 1.0$) estimation error overall.

2-2- Proposed method

Our proposed method starts with acquiring color and depth data using a Kinect V.2 sensor. Next we apply preprocessing techniques such as noise reduction, brightness adjustment to account for background color variability, and illumination respectively. Next, we apply thresholding based segmentation of color and depth images to extract the fruit object from the background. The smallest pixel value in depth images finds and system is used to normalize the distance between the fruit object and the sensor. Finally, for calculating the weight of the fruit that is being imaged, the (1) is used,

$$Fruit\ weight = \alpha * \beta * \delta / \theta \qquad (1)$$

Where, α is average of segmented color image dimensions size, β is distance from depth sensor in meter (converted from millimeter), δ is average weight of the respective fruit and θ is a constant (set to 100 for all the experiments reported here) for normalizing obtained final real number in to grams. For example, let us consider an object (sweet pepper fruit) in 1.3 meter distance from the sensor has 77 x 77 color image dimension size and its corresponding weight should be calculated. Using only the color image with wrong size leads to wrong weight estimation. The color image size should be normalized based on the distance between fruit and sensor before weight estimation. After calculating distance using depth sensor which is 1.3 meter, the process of weight estimation can be pursued since if an object is farther, the size is smaller, and vice versa. In the example case, mean of dimensions sizes, which is 77 multiplied to the calculated distance which is 1.3 and the result of each dimension will be 1000.1 millimeter which is close to 1.0 meter. After multiplying gained number to average weight of the fruit which is 75 grams, and dividing to 100, there is exist final weight of 75.07 grams. Actual weight of the fruit in this example case was 73.59 gram so 1.48 gram error appeared for this sample. Figure 4 illustrates proposed fruit weighting method using RGB-D images with this sweet pepper fruit example.

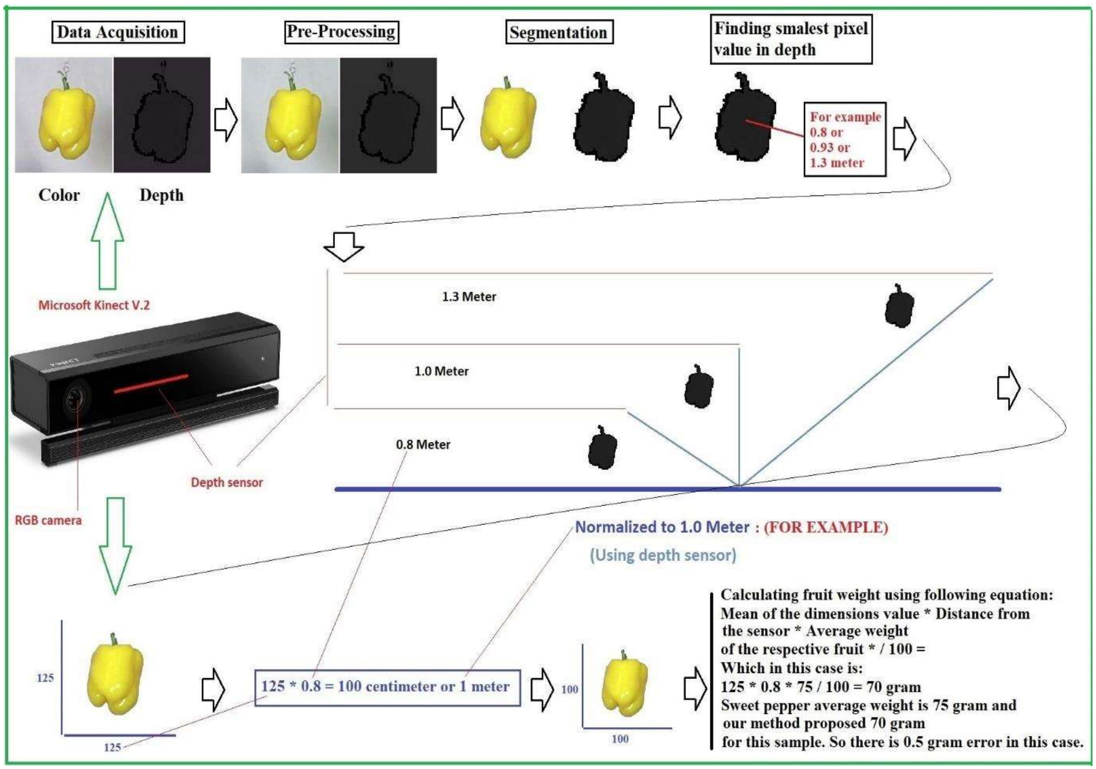

Figure 4 : Proposed fruit weighting method using RGB-D images with an example sweet pepper's weight estimation.

3. Experimental evaluation and results

In this section we provide the details of our dataset of sweet lemons, sweet peppers, and tomatoes due to the lack availability of open source alternatives. We further provide details of the imaging setup with Kinect V.2 depth sensor. Experimental results on the dataset along with different error metrics used for evaluation are presented.

3-1- Proposed RGB-D fruits database

Due to lack of RGB-D fruits datasets, especially for sweet lemon, sweet pepper and tomato, we have decided to create one. Our collected dataset consists of totally 150 samples with 50 samples for each fruits. As mentioned before the dataset is recorded using the Microsoft Kinect V.2 sensor which is capable of recording color (RGB) and depth images simultaneously. We used the following equipment and tools: white curtain background, stool, personal computer

(PC), tripod for mounting the sensor, Kinect V.2, objects (fruits), main light (spot-light), and back light (spot-light), see Figure 5. Average weight of the sweet lemon is 69 grams, average weight the sweet pepper and tomatoes is 75 and 62 grams respectively. Figure 6 represents some representative RGB-D fruit database samples showing the RGB and depth (shown as grayscale images).

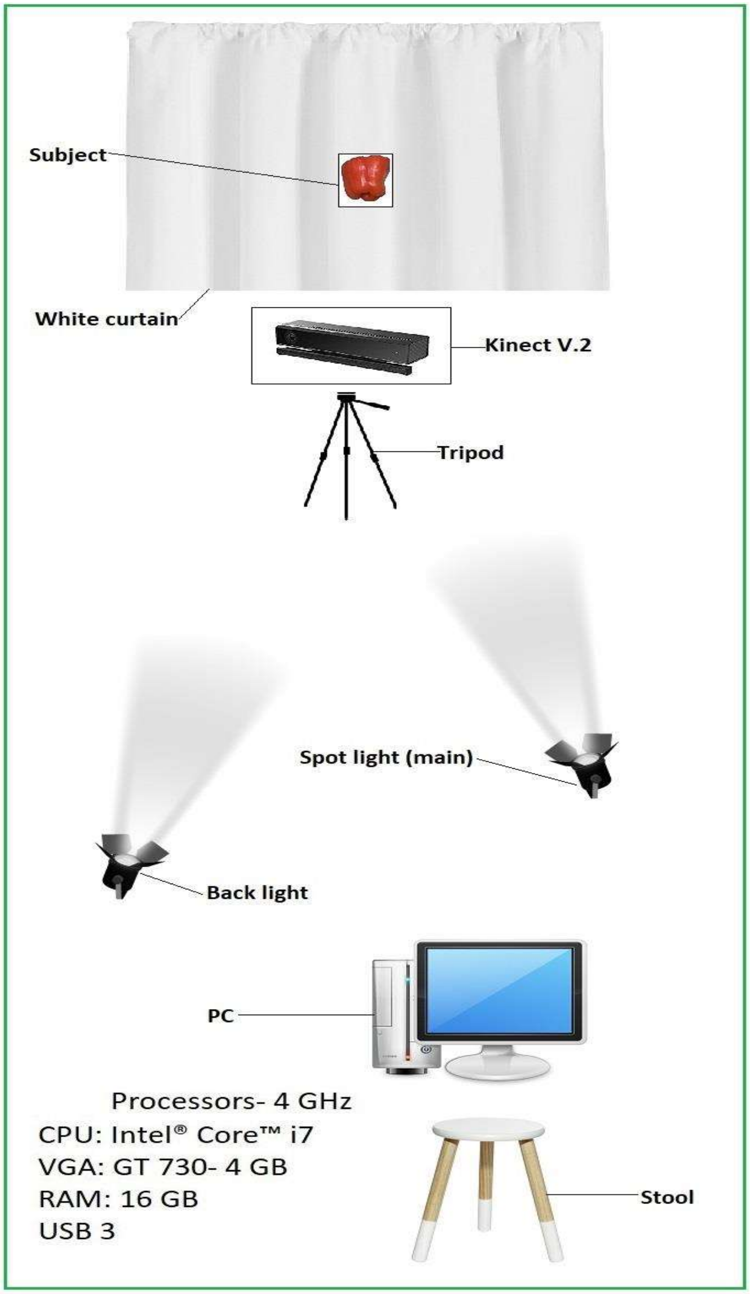

Figure 5 : Proposed RGB-D fruit databases recording setup

Figure 6 : Some samples from the proposed RGB-D fruits dataset. Sweet lemon (left in 0.8 meter), sweet pepper (middle in 1.0 meter) and tomato (right in 1.3 meter)

3-1- Error analysis

We utilize the following error metrics for estimating the automatic fruit weight estimation quantitatively.

- Mean absolute error (MAE): For evaluation of the RGB-D fruits database, Mean Absolute Error (MAE) [21] factor is employed. MAE is a measure of difference between two continuous variables. Assume X and Y are variables of paired observations that express the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. Consider a scatter plot of n points, where point i has coordinates (xi, yi)... Mean Absolute Error (MAE) is the average vertical distance between each point and the Y=X line, which is also known as the One-to-One line. MAE is also the average horizontal distance between each point and the Y=X line. The mean absolute error is a common measure of forecast error in time series analysis [22] where the terms "mean absolute deviation" is sometimes used in confusion with the more standard definition of mean absolute deviation. Table 1 represents proposed method result on proposed database with all 150 samples, validated using MAE factor.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n} \qquad (2)$$

- Mean squared error (MSE): This is an estimator (for estimating an unobserved quantity) measures the average of the squares of the errors or deviations, which is, the difference between the estimator and what is estimated. The MSE is a measure of the quality of an estimator. It is always non-negative, and values closer to zero are better.

$$\mathrm{MSE} = \frac{1}{\mathrm{M} \times \mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{N}-1} \sum_{\mathrm{j}=0}^{\mathrm{M}-1} [\mathrm{X}(\mathrm{i},\mathrm{j}) - \mathrm{Y}(\mathrm{i},\mathrm{j})]^2 \qquad (3)$$

In which, X and Y are two arrays with the size of M*N. To any extent Y resembles X, the value of MSE will reduce.

- Root mean square error (RMSE): This has the same units as the quantity being estimated. For an unbiased estimator, the RMSE is the square root of the variance, which is known as the standard deviation [23] [24].

- Mean Error (ME):
  Mean Error= mean (real weight- estimated weight) (5)

- Standard deviation error (STDE):

  STD Error= standard deviation (real weight – estimated weight) (6)

Table 2 presents evaluation result using MAE, MSE, RMSE, ME, and STDE quality metrics on all the samples at three different distances from the depth sensor. As it is clear in Table 2, the results are satisfactory and there are just 5.3 overall error for all 150 samples at the end for MAE, 7.32 for MSE and 2.70 for RMSE, 0.57 for ME, and 2.7 for STDE. From Table 2 we see that the best recognition belongs to sweet pepper with 4.5 error percentage in three different distances, sweet lemon and tomato are in second and third place with 5.4 and 6.2 error percentage respectively. Final overall error for all 150 samples is 5.3 error percentage or 94.7 recognition accuracy in MAE. Further, lowest error percentages for MSE and RMSE are belong to sweet pepper too (just like MAE), with 6.31 and 2.51 values respectively. Figure 7 shows actual versus estimated weight for 50 sweet pepper samples graphically. In this figure, x axis represents the number of sweet peppers and y axis represents the weight in gram. Figure 8 represents MSE, RMSE, ME, and STDE for the 50 samples of sweet lemon for all distances. Overall, our image processing approach obtained good results with less error in comparing with the ground-truth weights of these tested fruits. We believe this feasibility study can be extended to be deployed in a real scenario where the background variation need to be modelled and this defines an ongoing research in this work. Further, different distances and including extra depth sensors in a multi-directional fusion approach can improve the weight estimation accuracy with depth sensors.

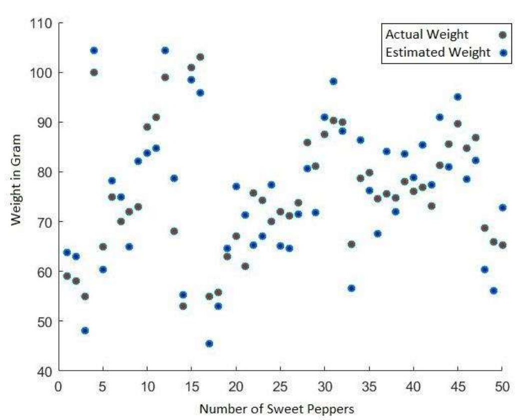

Figure 7: Actual versus estimated weight for 50 sweet pepper samples.

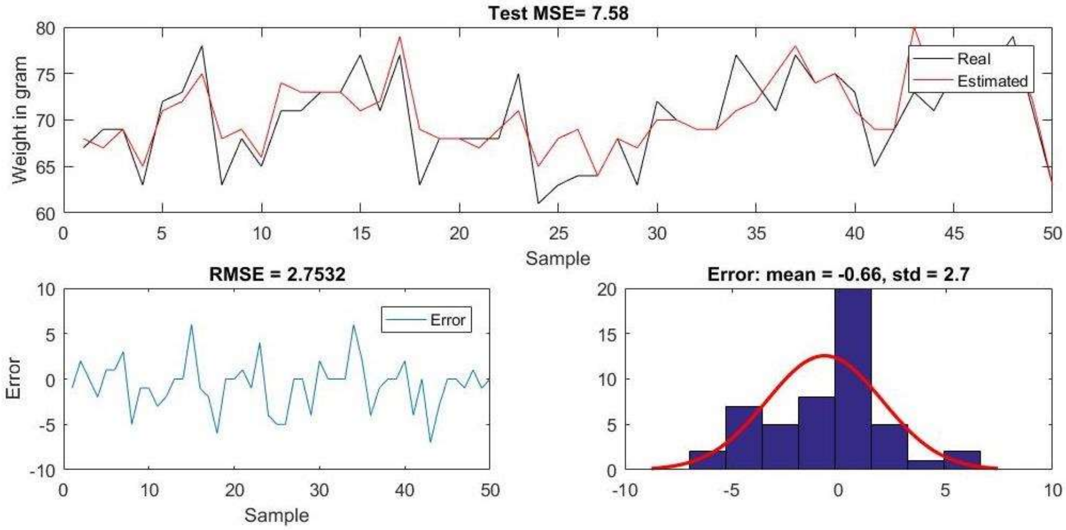

Figure 8: MSE, RMSE, Mean Error and STD Error belong to 50 samples of sweet lemon for all distances.

Table 2: Proposed fruit weight estimation results on RGB-D fruits database, evaluated by MAE, MSE, RMSE, ME, STDE error metrics for three different distances from the depth sensor.

-	MAE	MSE	RMSE	Mean Error	STD Error	Distance from Sensor
Sweet Lemon	5.9	7.87	2.80	0.59	1.9	0.8 Meter
Sweet Paper	4.7	6.47	2.54	0.47	2.4	0.8 Meter
Tomato	6.6	7.59	2.75	0.57	2.3	0.8 Meter
-	-					-
Sweet Lemon	4.4	7.34	2.70	0.60	2.8	1.0 Meter
Sweet Paper	2.8	6.11	2.47	0.36	3.1	1.0 Meter
Tomato	4.5	7.94	2.81	0.47	2.6	1.0 Meter
-	-					-
Sweet Lemon	6.0	7.55	2.74	0.79	3.5	1.3 Meter
Sweet Paper	6.1	6.35	2.51	0.70	3.0	1.3 Meter
Tomato	7.5	8.69	2.94	0.66	2.9	1.3 Meter
-	-					-
Sweet Lemon (All Distances)	5.4	7.58	2.75	0.66	2.7	0.8 – 1.0 – 1.3 Meter
Sweet Paper (All Distances)	4.5	6.31	2.51	0.51	2.8	0.8 – 1.0 – 1.3 Meter
Tomato (All Distances)	6.2	8.07	2.84	0.56	2.6	0.8 – 1.0 – 1.3 Meter
-	-					-
All 150 Samples	5.3	7.32	2.70	0.57	2.7	0.8 – 1.0 – 1.3 Meter

3. Conclusions

This paper proposed a method to calculate three fruits (sweet lemon, sweet pepper and tomato) weight using RGB-D images in three different distances of 0.8, 1.0 and 1.3 meter. Kinect V.2 sensor could work very efficient in this place as a cheap replacement for industrial machines. Using depth images beside to color images, could handle need to industrial robots and color image weakness to calculate distance between object and sensor. As achieved results are so promising and satisfactory with 94.7and 92.67 recognition accuracy for MAE and MSE factors, it is suggested to use this technique for other fruits or even objects (having average weight of the fruit or object). Also, it is suggested to use even stronger sensors for more than 5.0 meter recording distance capability and higher resolution in both RGB and depth sensors. It is possible to use this method on moving robots as well; however, it will require robust background modeling to avoid the depth noise. Using such automated techniques could lower the cost of human sources, and further decreases the industrial machine source costs as well and helps us to optimize agriculture industry.

References

[1] Automatic fruits weighting and grading, http:// food-machines.org/product/fruit-processing-machine/ fruit-sorting-machine.html, accessed: 2012-09-30.

[2] R. C. Gonzalez, R. E. Woods, Digital image processing (2012).

[3] P. Lucas, L. Van Der Gaag, Principles of expert systems, Addison-Wesley Wokingham, 1991.

[4] S. Ganguly, D. Bhattacharjee, M. Nasipuri, 3d image acquisition and analysis of range face images for registration and recognition, in: Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing, IGI Global, 2016, pp. 185-216.

[5] Computer arts / 3d world glossary, https://osu.pb.unizin.org/graphicshistory/back-matter/glossary/Documentretrieved26April2018., accessed: 2012-09-30.

[6] A. M. Pinto, P. Costa, A. P. Moreira, L. F. Rocha, G. Veiga, E. Moreira, Evaluation of depth sensors for robotic applications, in: Autonomous Robot Systems and Competitions (ICARSC), 2015 IEEE International Conference on, IEEE, 2015, pp. 139-143.

[7] H. Sarbolandi, D. Leoch, A. Kolb, Kinect range sensing: Structured-light versus time-of-ight kinect, Computer vision and image understanding 139 (2015) 1-20.

[8] L. Cruz, D. Lucio, L. Velho, Kinect and rgbd images: Challenges and applications, in: Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2012 25th SIBGRAPI Conference on, IEEE, 2012, pp. 36-49.

[9] R. A. El-laithy, J. Huang, M. Yeh, Study on the use of Microsoft Kinect for robotics applications, in: Position Location and Navi- gation Symposium (PLANS), 2012 IEEE/ION, IEEE, 2012, pp. 1280-1288.

[10] K. Aitpayev, J. Gaber, Creation of 3d human avatar using kinect, Asian Transactions on Fundamentals of Electronics, Communication & Multimedia 1 (5) (2012) 12-24.

[11] B. Y. Li, A. S. Mian, W. Liu, A. Krishna, Using kinect for face recognition under varying poses, expressions, illumination and disguise, in: Applications of Computer Vision (WACV), 2013 IEEE Workshop on, IEEE, 2013, pp. 186-192.

[12] J. Ding, C.-W. Chang, Feature design scheme for kinect-based dtw human gesture recognition, Multimedia Tools and Applications75 (16) (2016) 9669-9684.

[13] C. Chen, K. Liu, N. Kehtarnavaz, Real-time human action recognition based on depth motion maps, Journal of real-time image processing 12 (1) (2016) 155-163.

[14] S. J. Fluckiger, Security with visual understanding: Kinect human recognition capabilities applied in a home security system, Ph.D. thesis (2012).

[15] C. Teoh, A. M. Syaifudin, Image processing and analysis techniques for estimating weight of chokanan mangoes, Journal of Tropical Agriculture and Food Science 35 (1) (2007) 183.

[16] S. Nuske, S. Achar, T. Bates, S. Narasimhan, S. Singh, Yield estimation in vineyards by visual grape detection, in: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, IEEE, 2011, pp. 2352-2358.

[17] D. Font, M. Tresanchez, D. Mart__nez, J. Moreno, E. Clotet, J. Palac__n, Vineyard yield estimation based on the analysis of high resolution images obtained with arti_cial illumination at night, Sensors 15 (4) (2015) 8284-8301.

[18] M. Omid, M. Khojastehnazhand, A. Tabatabaeefar, Estimating volume and mass of citrus fruits by image processing technique, Journal of food Engineering 100 (2) (2010) 315{321.

[19] W. Spreer, J. Muller, Estimating the mass of mango fruit(mangifera indica, cv. chok anan) from its geometric dimensions by optical measurement, Computers and electronics in agriculture75 (1) (2011) 125-131.

[20] F. Marinello, A. Pezzuolo, D. Cillis, L. Sartori, Kinect 3d reconstruction for quanti_cation of grape bunches volume and mass, Engineering for Rural Development 15 (2016) 876-881.

[21] C. J. Willmott, K. Matsuura, Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance, Climate research 30 (1) (2005) 79-82.

[22] R. J. Hyndman, A. B. Koehler, Another look at measures of forecast accuracy, International journal of forecasting 22 (4) (2006) 679-688.

[23] E. L. Lehmann, G. Casella, Theory of point estimation, Springer Science & Business Media, 2006.

[24] R. J. Hyndman, A. B. Koehler, Another look at measures of forecast accuracy, International journal of forecasting 22 (4) (2006) 679688.