Applying Linear Regression to Estimate Weight of Non Axi-Symmetric fruit

Hurriyatul Fitriyah[1], Eko Setyawan[2], Muhammad Rifqi Radifan Masruri[3]

[1,2,3]Computer Engineering Department, Faculty of Computer Science, Universitas Brawijaya
{[1]hfitriyah@ub.ac.id, [2]ekosetiawan@ub.ac.id, [3]rifqiradifan@gmail.com}

Abstract. Weight is an important parameter in fruits' quality identification. Measuring fruits' weight using scale is tedious since fruits must be taken from tree and placed on contact to scale. Many researches have proposed non-contact estimation methods of fruits' weight using 2D images. The studies were commonly applied in axi-symmetric fruits, such oranges. In this paper, an algorithm to estimate weight of non axi-symmetric fruit is developed. It used a Linear Regression rather than geometric-based methods as proposed by other researches. The non axi-symmetric fruits chosen was star fruits. It is a challenging fruits since its basic shape is not round but irregular star shape. The estimation used pixel count from one-view image of the fruits' projection as feature. The proposed method has RMSE of 16.322 Gram and MAPE of 7.089% compare to the expected weights. It also has high Coefficient of Determination, R^2, 0.8829 compare to the weight scale measurement.

Keyword : Weigth, Fruits, Regression

1 Introduction

Regression is one of basic learning method for prediction. It learns data trends from previous data that is named data training. The trend between dependent variable and independent variables is shown as linear and non-linear line where its coefficients was found based on data training. The coefficients found can be used to predict outcome of future data. This learning method is very powerful and has been successfully used to predict physical properties such height and weight of objects based on visual measures. Research by [1] used Linear Regression to estimate adults' height and mid-arm circumference. Research by [2] used Linear Regression to predict body weight of dairy cows based on 3D visual data. Other research by [3] develop an algorithm to estimate volume of shoulder muscle based on cross-sectional area using Linear Regression. Those researches promotes a non-direct measurement of objects' properties i.e. height, weight, volume, based on other related information.

Fruits is also one of objects where its physical properties are also important to be measured. Marketing standard of fruits usually includes measure of Maximum Diameter of the Equatorial Section (MDES), Fruit Weight (FW) and Circumference (C). This paper is focused on Fruit Weight (FW) as it is cannot be seen visually, compare to MDES and C. In EU, fruits must met specific weight in order to be marketed for example as stated in EU marketing standard for fruits and vegetables [4]. In Indonesia, Grading system to market fruits such apple [5] and mango [6] also used weight as one of its parameters. Fruits' weight is commonly measured by weighting

scale or load sensor. This measurement require fruits to be placed in contact with the tools. A non-contact measurements are more preferred as it won't degrade the shape, size and condition of fruits which are soft and succulent.

Computer vision has been widely used for this particular type of measurement. The measurement enable measurement of fruits in factory while the fruits are running in conveyor belt or even in field without picking out the fruits from trees. Many researches in estimation of fruits' weight using 2D images used Regression as learning method to estimate fruits' weight from its pre-determined volume. It has shown good accuracy in estimating weight of apple, sweet-lime, lemon and orange [7] with $R^2 > 0.86$ and macaw palm fruits [8] with R^2 of 0.837. Other research estimated weight of apples using Linear Regression based on its diameter and area and show accuracy of 96.5% [9]. Other research estimate mango weight using Linear Regression based on counted pixel of mango projection in 2D images and show good result of R^2 0.9769 and percentage error of 3.76% [10].

Most weight estimation of fruits that was based on 2D image were developed for round, spherical or axi-symmetrical fruits, e.g. lemon, apple, and mango. To our acknowledgement, non-axi symmetrical fruits, for example star fruit, has not been explored yet. This fruit has challenged since the shape is not round or spherical hence basic shape volume equation cannot be applied to estimate its volume that lead to weight estimation. Star fruit, or in Latin Averrhoa Carambola, is a very juicy fruits that is mostly cultivated in South-East Asia. It taste sweet and sour. This tropical fruit has a very unique shape where it has 5 distinctive ridge. The 5 ridges result a star-shape when cut in cross-section thus it is famous as star fruit.

This study proposed a black box method where the relation between star fruits' visual information and its weight was not modeled. The relation is trained using Linear Regression using several data training and tested on data testing.

2   Method

An algorithm to estimate weight of star fruits was developed in this study. The algorithm consist of 4 steps which are: (1) Image acquisition, (2) Image Processing, (3) Feature Extraction, (4) Weight Estimation. In general, the algorithm utilizes only one-view image of star fruit to estimate its weight based on pixel count of the fruit's projection. The weight estimation used Linear Regression.

2.1   Image Acquisition

Star fruits with different sizes were photographed in a mini studio. Background, lighting and distance between fruit and camera were fixed in order to simplify segmentation process. Background was set to white as it has the lowest saturation, hence segmenting the fruits were easy. Lighting is fixed in every acquisitions to maintain saturation value of all 2D images. Distance between object and camera is fixed to 45 cm to maintain pixel-to-cm ratio. The set up for image acquisition is shown in Figure 1.

Star fruits has 5 ridges hence 5 different views were photographed. The star fruits stand up in its two ridges where its stalk was placed horizontally. It was rotated to yield 5 multi-views of single fruit. Star fruits are natural products hence it has slight irregularity in length among the ridges. Weight of a star fruit was measured based on a single-view as if the fruits were measured in trees. The multi-views would be used to investigate variation between views. Example of 5-views from a single star fruit is shown in Figure 2. As can be seen, the projections was slightly vary between views.
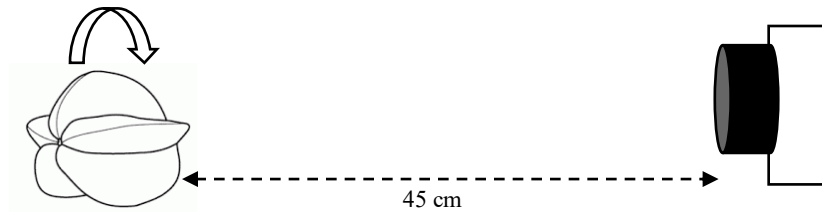
Fig. 1. Set-up for data acquisition

Fig. 2. Example of 5 different views taken from a single star fruit.

2.2   Image Processing

Several image processing steps were performed in order to segment the star fruits from its background. The background was made easy which is white as it has very low saturation compare to other color. The similar result can also be yielded when the segmentation was performed in RGB channel since white color has very high value of Red, Green and Blue. The Saturation was chosen as the segmentation only need to be performed in one channel. The RGB conversion to Saturation used Equation 1.

$$S = \begin{cases} 0 & , C_{max} = 0 \\ \frac{C_{max} - C_{min}}{C_{max}} & , C_{max} \neq 0 \end{cases} \tag{1}$$

$$\text{where } C_{max} = \max\left(\frac{R}{255}, \frac{G}{255}, \frac{B}{255}\right) \text{ and } C_{min} = \min\left(\frac{R}{255}, \frac{G}{255}, \frac{B}{255}\right)$$

The image in Saturation color-space was then segmented using the famous Otsu's Thresholding method. The Otsu's thresholding method found a threshold that would separate the data into 2 classes, background and foreground. The threshold was chosen as a value that gives the highest inter-class variation and at the same time gives the lowest intra-class variation of 2 classes. The Segmented Image has binary values where 1 is for the pixels belongs to star fruits and 0 to the white background. The image processing step was shown in Figure 3, whilst the result on a star fruit as an example is shown in Figure 4. The segmentation process based on Saturation color-space shows a very good segmentation result.
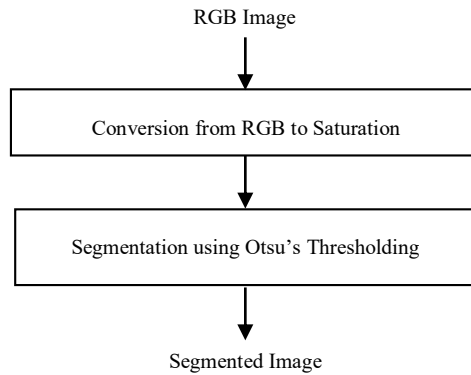
RGB Image

| Conversion from RGB to Saturation |

| Segmentation using Otsu's Thresholding |

Segmented Image

Fig. 3. Image processing steps

(a)

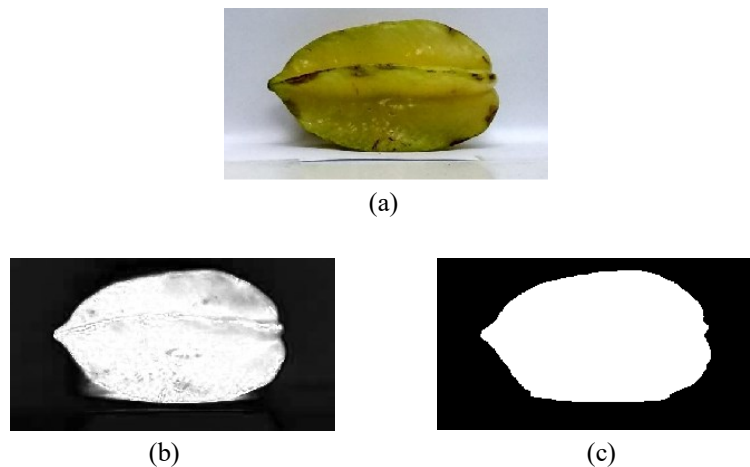(b)   (c)

Fig. 4. Result of Image Processing steps to segment star fruit; (a) RGB image, (b) Saturation Image, (c) Segmented Image

2.3   Feature Extraction

Several features that related to size can be extracted from those Segmented Images such as size of bounding box (height and width) and pixel count. Illustration of the two features are given in Figure 5. This paper tried to use only one feature for prediction, hence a correlation between each features and fruits' weight was tested. The two features was area of the star fruits, which are: (1) area of bounding box, height × width; (2) area of star fruit projection or pixel count. The area of bounding box represents the maximum area of fruits' projection, whilst pixel count represents cross-section area.

The squared Correlation Coefficient, R^2, for area of bounding box to the weight is 0.747 whilst pixel count or area of projection is 0.7795. The data and its squared Correlation Coefficient is shown in Figure 6. Although both coefficients are nearly similar, but the pixel count has higher value. Hence pixel count is chosen as feature for prediction.
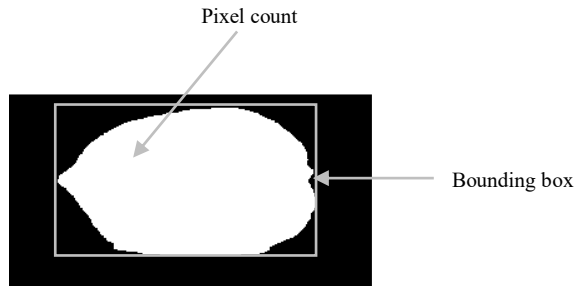

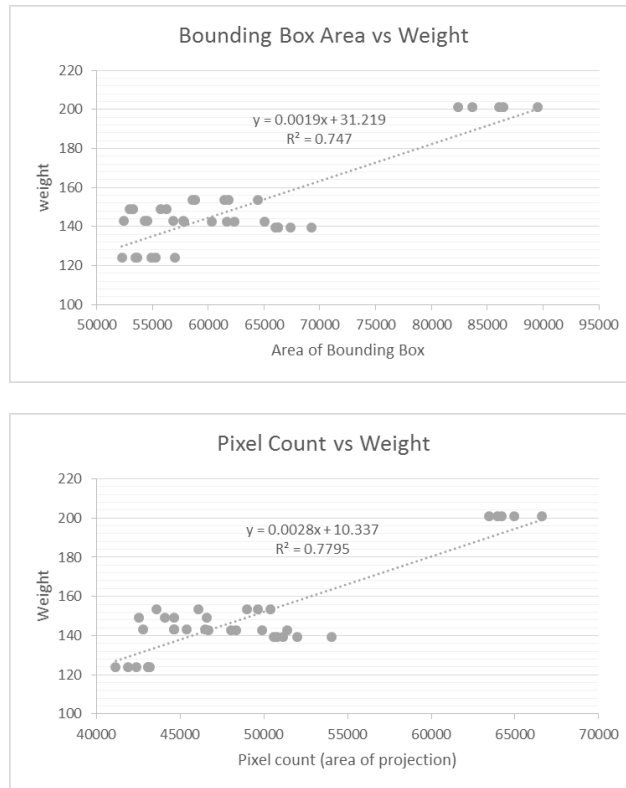Fig. 5. Illustration of area based on bounding box and pixel count

Fig. 6. Data and its squared correlation coefficient, R^2 for: (a) Bounding box area, (b) Pixel count

2.4   Weight Measurement

Regression is a basic method used for prediction. It has several type of Regression e.g. Linear, Logarithmic, Exponential. Decision to use which type of Regression is mostly based on plot of data. As shown in Figure 4 (b), pixel count has a high linear correlation toward the weight. Hence this study used the simple Linear Regression to estimate star fruits' weight which is based on the pixel count. Linear Regression determine a linear equation of $y = ax + b$ based on data training. It is basically an equation of a line that has gradient and y-axis cut-off. It has 2 coefficients, a and b, which were determined using Equation 2 and Equation 3, respectively, where x is pixel count of projection area and y is star fruits' weight.

$$a = \frac{N \sum_{n=1}^{N} x_n y_n - \left(\sum_{n=1}^{N} x_n\right)\left(\sum_{n=1}^{N} y_n\right)}{N \sum_{n=1}^{N} x_n^2 - \left(\sum_{n=1}^{N} x_n\right)^2} \qquad (2)$$

$$b = \bar{y} - a\bar{x} = \frac{\sum_{n=1}^{N} y_n}{N} - m \frac{\sum_{n=1}^{N} x_n}{N} \qquad (3)$$

3   Result and Analysis

A total of 35 images were collected from 7 star fruits with 5 multi-views. The star fruits has various weight ranging from 124 Gram to 201 Gram that were measured using electronic weight scale. Sample of images from 7 star fruits are shown in Figure 7. The weight information is given in each image. As shown in Figure 7, some star fruits has different length among the ridges. Thus when rotated, it would give a slight variation on its projection in the 2D images.
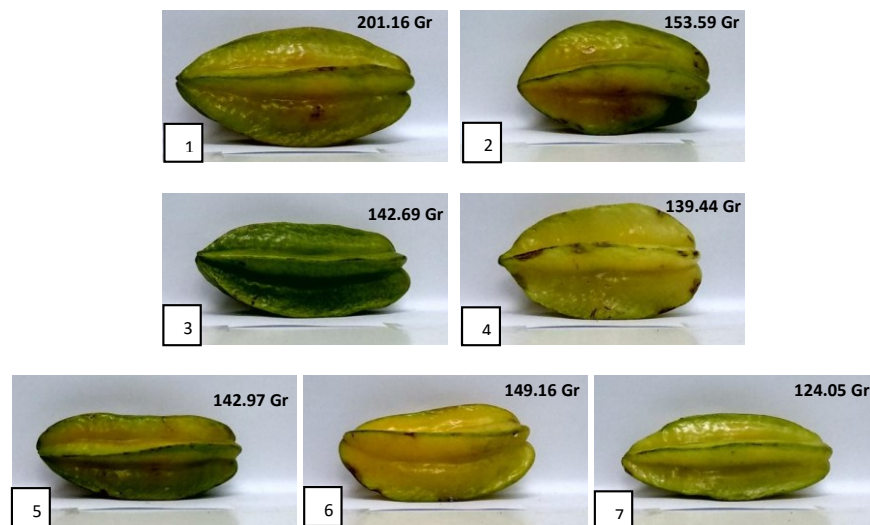

Fig. 7. Images taken and its weight

3.1 Variation of Pixel Count among Views

Each star fruit is photographed 5 times to simulate 5 possibilities to view the fruit. The pixel count was determined and analyzed in term of variation. Range, Standard deviation (SD), Mean and Coefficient of Variation (CV) is shown in Table 1. CV is calculated as $SD/mean$. General rule for CV mentioned that if the value below 1 (CV<1) then the data variation is low. The Table 1 shows that all CV of fruits are very low, meaning that pixel count is not vary much between views. The Range are also small compare to the Mean of pixel count. Hence pixel count can be used to estimate weight regardless of the view when photographing the star fruits. The Range, Standard deviation and Mean are in Gram.

Table 1. Variation of pixel count among 5-views of star fruits

Fruit	Range	Standard Deviation	Mean	Coefficient of Variation
1	3133	1217.5	64640.6	0.019
2	6797	2829.2	47727.2	0.059
3	4730	1821.1	48867.6	0.037
4	3474	1420.6	51711.4	0.027
5	3683	1345.4	44818.75	0.030
6	4026	1476.4	44270	0.033
7	2037	815.3	42338.8	0.019

3.1 Accuracy of Weight Estimation

The actual weights were compared to the estimated weights using RMSE (Root Mean Squared Error), MAPE (Mean of Absolute Percentage Error) and squared coefficient of correlation, R^2. The squared Correlation of Coefficient is also known as coefficient of determination. Seven-fold cross validation was used to determine overall RMSE and MAPE. From 7 star fruits, 6 fruits were used as data training against 1 fruits for data testing, alternately. The result is shown in Table 2 where RMSE are in Gram.

Table 2. RMSE and MAPE of algorithm in 7-folds cross validation

	Fold							Mean
	1	2	3	4	5	6	7	
RMSE	41.418	12.026	8.533	21.185	8.007	14.234	8.855	16.322
% error	12.543	5.337	4.761	9.725	4.765	6.766	5.726	7.089

The Table 2 shows that the developed algorithm was able to estimate weight of star fruits with MAPE of 7.089%. The average MAPE between actual and estimated weight are also quite low which is 16.322 Gram. The coefficient of determination, R^2, of actual and estimated weight are also high, which is 0.8829. Thus the proposed algorithm that

use only single view of star fruits' 2D images has high correlation compare to weighting scale to measure weight of star fruits.

4   Conclusion

This research has developed a method to estimate weight of star fruits using only one-view image. The method estimate weight using Linear Regression based on pixel count of the fruits' projection. Pixel count has low variation amongst views, thus it is good feature to be applied in real-condition where camera can only capture one-view of the fruits in a single acquisition. The proposed method has shown to have a low RMSE and MAPE compare to the actual weight. The method also has high correlation with weight measurement that use weight scale. Future research should test the proposed method on other non axi-symmetrical fruits. This is to test whether the physical properties of fruits as natural product, which has irregularity in shape, still can be estimated using a simple Linear Regression.

Acknowledgement

References

1. Y. Liu, A. Sowmya and H. Khamis, "Single camera multi-view anthropometric measurement of human height and mid-upper arm circumference using linear regression," *PLoS One,* vol. 13, no. 4. (2018).

2. X. Song, E. Bokkers, P. Van der Tol, P. Groot Koerkamp and S. van Mourik, "Automated body weight prediction of dairy cows using 3-dimensional vision," *Journal of Dairy Science,* vol. 101, no. 5, pp. 4448-4459. (2018).

3. H. Henninger, G. Christensen, C. Taylor, J. Kawakami, B. Hillyard, R. Tashjian and P. Chalmer, "The Muscle Cross-sectional Area on MRI of the Shoulder Can Predict Muscle Volume: An MRI Study in Cadavers," *Clinnical Orthopaedic and Related Research".* (2019).

4. R. P. Agency, "Guidance: Marketing standards for fresh fruit and vegetables," 4 October 2018. [Online]. Available: https://www.gov.uk/guidance/comply-with-marketing-standards-for-fresh-fruit-and-vegetables. [Accessed 19 November 2019].

5. "SNI 8024:2014 Apel," Badan Standardisasi Nasional Indonesia, Jakarta. (2014).

6. "SNI 3164:2009," Badan Standardisasi Nasional Indonesia, Jakarta. (2009).

7. G. Venkatesh, S. Iqbal, A. Gopal and D. Ganesan, "Estimation of Volume and Mass of Axi-Symmetric Fruits Using Image Processing Technique," *International Journal of Food Properties,* vol. 18, pp. 608-626. (2014).

8. A. Costa, R. Elisângela, A. Braga and F. Pinto, "Measurement of volume of macaw palm fruit using traditional and the digital Moiré techniques," *Revista Brasileira de Engenharia Agrícola e Ambiental,* vol. 20, no. 2, pp. 152-157. (2016).

9. O. Comert, M. Hekim and k. Adem, "Weight and Diameter Estimation Using Image Processing and Machine Learning Techniques on Apple Images," *International Journal of Engineering Research and,* vol. 9, no. 3. (2017).

10.   C. Teoh and A. Syaifudin, "Image processing and analysis techniques for estimating weight of chokanan mangoes," *Journal of Tropical Agriculture and Food Sciences,* vol. 35, no. 1, pp. 183-190. (2007).