

Multiple Food or Non-Food Detection in Single Tray Box Image using Fraction of Pixel Segmentation for Developing Smart Nutrition Box Prototype

Yuita Arum Sari, Jaya Mahar Maligan, SigitAdinugroho, Yusuf Gladiensyah Bihanda

Abstract: *Smart Nutrition Box is a hardware prototype to predict food leftover measurement as well as the nutrition of the leftover. In the previous approach, there was a need of trained observer to conduct the analysis. Human observer may produce subjective judgement, so that algorithm which is embedded in a prototype is proposed to get rid of the bias. Black background of tray box is used, and two menus are served in this paper. The problem on raw dataset of images is reflection and it affects the result of segmentation, since it is considered to determine the leftover measurement precisely. In this paper, we focus on how to classify image of food and non-food image in each compartment of tray box by using pixel segmentation before going to further stage of prediction. Automatic cropping is applied by means of rectangle contour detection for each compartment. Combination of L of HSL and V color channel of HSV color spaces are utilized to remove glare in each compartment. The ratio of segmented pixel is a fraction of detected object and the area of compartment. There are 10 out of 12 of tray box images containing multiple food that are correctly classified as food and non-food. The accuracy reaches 95.83% in all compartments using luminosity (L) 50% of lower upper white masking and 100% of upper white masking. It is proved that fraction pixel segmentation is sufficient to be embedded as one of features in Smart Nutrition Box prototype.*

Keywords: *food detection, food or non-food classification, image segmentation, contour detection*

I. INTRODUCTION

The development of applications for monitoring patient diets, especially in terms of knowing the remaining food calories are still under development [1]. This application contains an automatic prediction algorithm that can help nutritionists in estimating leftovers objectively and accurately. Some researchers have conducted research related to the prediction of nutrients from food [2][3]. In this paper we build a prototype which helps people to estimate

their leftover, called Smart Nutrition Box. The goal of this prototype is to detect what type of food and how many calories there are in the food automatically. The type of food used in this study is solid food placed in a tray box. Before understanding certain types of food and calorie content, it is necessary to detect how many types of foods are available in the tray box. Thus, the detection of whether a food exists in a tray box is proposed in this study.

To detect the presence of food in the compartment of a tray box, then the preprocessing stage needs to be conducted. In this case the preprocessing involved is segmentation. Segmentation is an important process in preprocessing to get useful information which in this case is the presence of food objects or not [4]. This segmentation is an algorithm that is still being developed to optimize the performance of the prediction of leftovers in the Smart Nutrition Box Prototype. Segmentation is widely applied for food recognition purposes. This segmentation becomes important especially for data taken with lighting that produces light reflection on the object to be segmented.

Previous studies applied improved segmentation of lighting results that cause reflections with tomato objects. With color channel based segmentation with V-Otsu it can produce less than 1% error when the objects are grouped [5]. In addition, further developments with the same object have been tested using a combination of two types of YUV and YCbCr color spaces. From the results of this study, the results of segmentation can reduce the error of about 3% by using several different types of cameras [6].

Because of the above reasons, experiments using color channels are still relevant to lighting condition and human vision perceives when taking pictures are carried out [7]. HSV and HSL is one color channel that is able to separate color from luminosity [8]. The purpose of the combination of the two types of color space is to improve the quality of segmentation caused by the reflection of light produced by the original object when photographed. This segmentation will also affect the quality of the proposed fraction of pixel segmentation method, so that multiple food or non-food detection is observed in each single tray box.

Revised Manuscript Received on January 5, 2020.

* Correspondence Author

Yuita Arum Sari, Computer Vision Research Group, Faculty of Computer Science, University of Brawijaya, Indonesia.

E-mail : yuita@ub.ac.id

Jaya Mahar Maligan, Food Nutrition Program, Faculty of Agricultural Technology, University of Brawijaya, Indonesia.

E-mail : maharajay@gmail.com

Sigit Adinugroho, Computer Vision Research Group, Faculty of Computer Science, University of Brawijaya, Indonesia.

E-mail : sigit.adinu@ub.ac.id

Yusuf Gladiensyah Bihanda, Informatics Engineering Program, Faculty of Computer Science, University of Brawijaya, Indonesia.

E-mail : chanvusufsj@student.ub.ac.id

II. MATERIALS AND METHOD

A. Dataset

There are 12 tray box images under black background taken from Smart Nutrition Box. Each tray box has four compartments. Based on that case, the total of compartment is $12 \times 4 = 48$ images of compartment. There are 2 menus of tray box to be applied in this case as depicted in the Figure 1. Each compartment may contain food or not, so that this detection is important for predicting leftover estimation afterwards.



Figure 1. Two menus in the tray box images

B. Research Methodology

The whole phase of the proposed method is displayed in the Figure 2. Smart Nutrition Box is completed with fixed camera on the top of tray box. Tray box is captured based on perpendicular position. The lighting condition is uniformly adjusted, then the captured image of tray box is saved in the storage. Building benchmark is utilized in order to make cropping result from each compartment in tray box has same size and this phase can be called as size calibration. The detail of this phase is presented in the following subsection. A tray box image have four compartments, and each of them is processed in the further stage as image segmentation. By the result of image segmentation, we applied our proposed method of pixel fraction to get the boundary between food and non-food image.

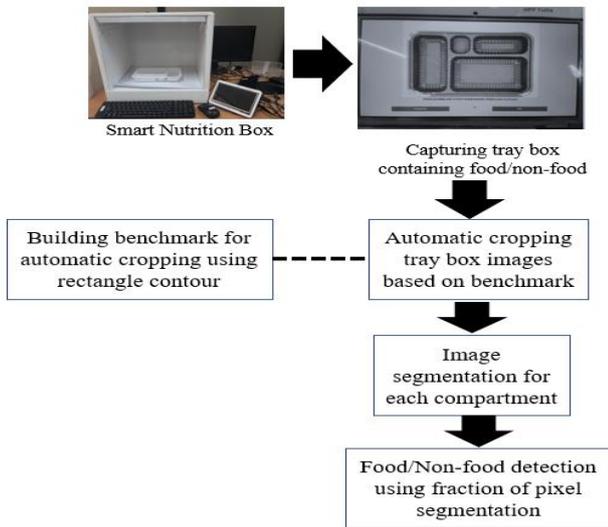


Figure 2. General phase of multiple food or non-food detection

a. Automatic Cropping Tray Box Image into Compartments

The first stage is cropping both benchmark and original image to have same dimension of image. This cropping in each compartments is also applied to enhance the quality of image segmentation result and predict the value of pixel ratio between detected object and the whole area of compartment.

Figure 3(a) represents the benchmark manually created by user regarding to the original size of compartment. Manual benchmark is built by drawing a line as the boundary in each compartment. As standart benchmark, we use black color as background and white line is signed as edge of compartment. This white line is determined as position in x and y coordinates and also return the value of width and size of each detected rectangle area. Figure 3 (b) shows the boundary is already detected as green line and 4 objects are recognized. Based on this result, the coordinates of each compartment are applied in the original image which the size of original image is suited as benchmark's size. Figure 4 presents the detail process of automatic cropping tray box into compartments.

The first stage is reading the benchmark image, then color transformation is computed since RGB is default. RGB is converted in grayscale image, then filtering is also applied since there still some noises, especially comes out in making line around boundary of compartment. In this filtering, we use 15×15 kernel size. After that, calculating a contour among the area of traybox is conducted. In this case, since the area of compartment near a rectangle shape, thus we count the number of contour approximation which has four sides. From contour detection, it retrieves the position of each rectangle object detection and returns those parameters to be applied in the original image. The result of automatic cropping based on rectangle contour detection is displayed in Figure 5. There are four compartments created with automatic cropping process.

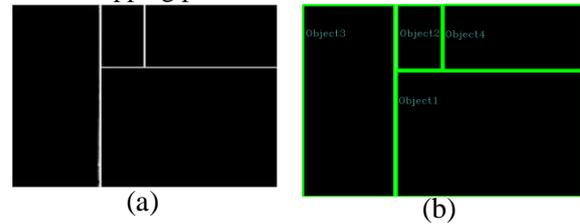


Figure 3. Manual benchmark (a) and detecting the area from each compartment using rectangle contour (b)



Figure 4. Finding rectangle contour in benchmark image

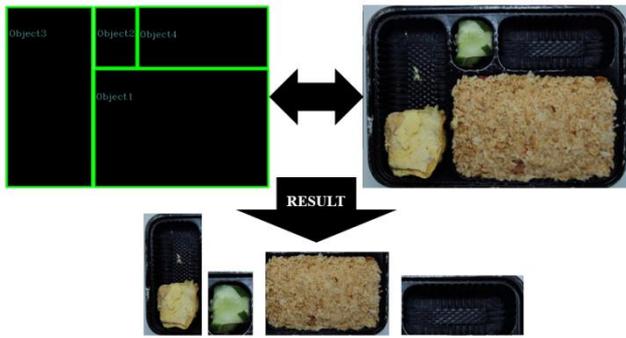


Figure 5. Result of automatic cropping into compartments for a tray box image

b. Improved Segmentation for Each Compartment

Segmentation is essential stage to have good detection of food or non food in each compartment. However, when taking raw dataset, the light reflection occurs and it will affect wrong segmentation. For instance, the object of non food can be determined as food since there is a glare in area of compartment. In this paper, we use two stage of segmentation to reduce some noises caused by the reflection. Figure 6 tells the whole process of segmentation process.

The first stage segmentation is applied to get the binary image by means of Otsu thresholding. The original image is transformed into HSV color space and only V color channel is taken. V color channel in HSV represents as Value or luminosity of image which has range from 0-100%. We take this channel since it influences significantly when taking raw picture of traybox image. Then, in order to reduce some unused pixel called as noises, we implement Gaussian filtering using 15x15 of kernel size. After that, ordinary Otsu thresholding is applied to obtain the binary image. The result of binary image is forwarded to 2nd phase of segmentation.

In the second stage of segmentation, morphology is used to get the focus area of segmented image. In this case, we use dilation and erosion consecutively using ellipse structuring element. Because of food object is not rigid as rectangle image, the most suitable structuring element is ellipse, since it can affect the smoothing process after erodes the boundary among pixels. To strengthen the object then median filtering is calculated using 15x15 of kernel size. After having masking binary image then convert it back into RGB color spaces.

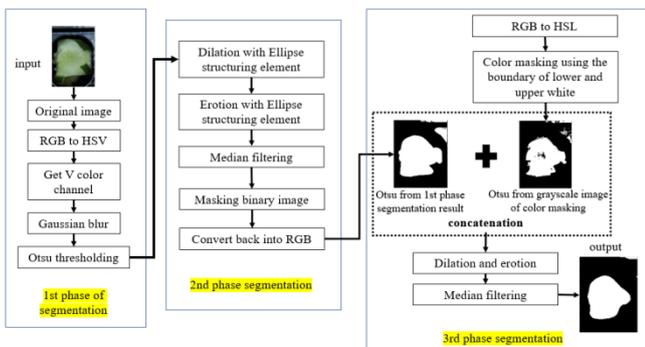


Figure 6. Three phases of segmentation process

These segmentation result of the second phase is processed to 3rd part of segmentation. There are two types of segmentation phases, first, from the result of second phase of segmentation is converted to binary image using Otsu thresholding and second from the same result of second phase of segmentation

is converted into HSL. Lighting color channel is set to be processed rather than other color channels of HSL. From L color channel we can detect glare information by using the boundary of lower and upper white color. By default, upper white color is always 100%. While, in this paper, we observe 45% and 50% as lower white boundary. Figure 7 shows the difference color between upper and lower white by using RGB color space. From both ordinary segmentation using Otsu dan L channel Otsu thresholding, then both of image are concatenated to complete the main food object using AND operator. The next step is dilation and erosion with 3x3 of kernel size. The last step is filtering to remove remaining noise using median filtering with 25x25 of kernel size. Figure 8 shows the result of each phase of segmentation.

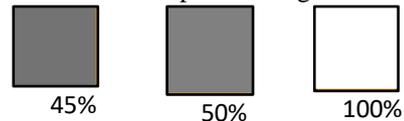


Figure 7. Lower white is 45% and 50%, upper white is 100% (presented in RGB color space)

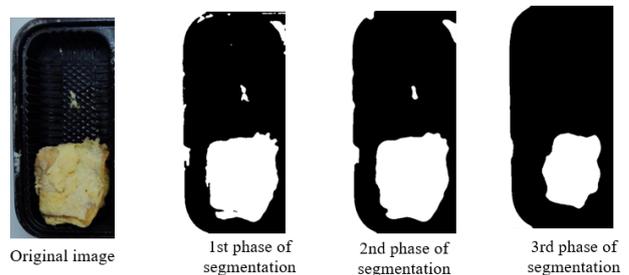


Figure 8. Preprocessing result phase in three phases

c. Fraction of Pixel Segmentation for Food or Non-Food Detection

From the final segmentation results in the previous three stages, the next step is to determine which areas are detected as food or not. Equation (1) represents formula to calculate the ratio between non zero pixel of binary image and the whole size of it.

$$FoS = \frac{count(px_0)}{n \times m} \quad (1)$$

where *FoS* is fraction of pixel segmentation, *px₀* is pixel which has non-zero value of binary image, while *n* × *m* is the width and height of segmented image from each compartment. Figure 9 illustrates the detected object based on black areas. Later, Equation (2) is applied to analyze which areas are detected as food and which are not in each compartment.

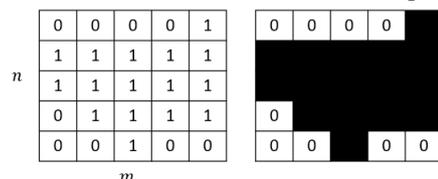


Figure 9. Counting *FoS*

$$food_{detect} = \begin{cases} food & , FoS > threshold \\ non - food & , otherwise \end{cases} \quad (2)$$

In this paper, we use *threshold* greater than 0.9% of fraction pixel of segmentation to be acknowledged as food. It means the detected area of non-zero value almost detected in area of compartment. The threshold can be changed according to the user preference. However, giving input parameter of threshold less than 1% is preferable in this case, since it fits as optimal *threshold* for given segmentation result. Therefore, segmentation phase will be determined the quality of image and detection precisely. As an example of food or non food detection by using *FoS*, from Figure 10, there are 15 non-zero pixel then the size of dimension of *nxm* is 25, so that the *FoS* will be 15 divided by 25 is 0.6 which is greater than *threshold* of 0.9%. By using Equation (2) we can conclude that in Figure 9 is recognized as food.

III. RESULT AND DISCUSSION

A. Food or Non-Food Detection Result in Each Tray Box

In this paper, two luminosity of HSL values are treated, 45% and 50% as stated in the subsection 2.2.2. There are 12 tray box images that has two menus. Figure 10 and 11 shows example of dataset that is correctly detected as food or non food in each compartment of tray box images using L of 45% and 50% of luminosity of HSL, respectively.



Figure 10. Correct food or non food detection using L 45%



Figure 11. Correct food or non food detection using L 50%

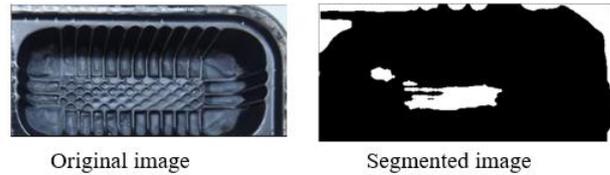


Figure 12. Wrong detection because of segmentation result

Ten out of 12 images of traybox are correctly classified as food or non-food either using 45% or 50% of L in HSL. Two tray box incorrectly detected since the result of segmentation phase is fail. So the area of *FoS* also wrong calculated. Table 1 shows two tray box images which has wrong detection in some compartments and its explanation. Based on Table 1, the experiment using luminosity 50% is much better than using 45% in detecting food or non-food in compartment. The result shows that using luminosity 50% gives 95.83% of accuracy, whilst 45% is 93.75%. One drawback of this method is still of finishing the glare problem, although this algorithm run well in almost dataset. The segmentation problem comes out when compartment is blank with many reflection surrounding the compartment. Another problem also happened when the background of compartment is textured and totally blank inside. It makes more noises in having reflection such as described in the Figure 12. The wrong detection mostly caused by blank compartment, although we already applied multiple smoothing in every stage of segmentation process.

Table 1. Wrong food or non-food detection.

Luminosity	Wrong detection	Note
HSL 45%		Left side: Two compartments are wrong classified. It should be non-food, but detected as food. Right side: A compartment is wrong classified.
HSL 50%		Left side: one compartment is wrong detected as food. It should be non food. Right side: also has the same wrong classified.

IV. CONCLUSION

In this paper, fraction of pixel segmentation algorithm is proposed to enhance the quality of prediction feature in Smart Nutrition Box. The triple segmentation phase achieves good to be calculated in fraction of segmentation method in order to classify food or non food in each compartment. It is proved by the result of detection reach 83.3% of accuracy from whole tray box images success to correctly detected food or non-food in both of 45% and 50% of lower white of luminosity in HSL color space. Based on compartment images, it achieves 95.83% for lower white of HSL 50% while 45% reaches 93.75% of accuracy. This algorithm always success when there is food in the compartment. However, when the compartment is empty and textured, it is a slightly more improvement in segmentation phase. For future works, the compartment which is detected as food will be recognized as one particular type of food.

ACKNOWLEDGMENT

The authors would like say thanks to students and research assistant who really helpful to have discuss to build and develop our current prototype, Smart Nutrition Box. Thank you as well to Brawijaya International Conference on Multidisciplinary Sciences and Technology for providing the facility to disseminate the research in wider scope and area.

REFERENCES

1. Ege, T., & Yanai, K. (2017, October). Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017 (pp. 367-375). ACM.
2. Akhi, A. B., Akter, F., Khatun, T., & Uddin, M. S. (2018). Recognition and Classification of Fast Food Images. *Global Journal of Computer Science and Technology*.
3. Kohila, R., & Meenakumari, R. (2017, March). Predicting calorific value for mixed food using image processing. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS) (pp. 1-4). IEEE.
4. Inunganbi, S., Seal, A., & Khanna, P. (2018, March). Classification of food images through interactive image segmentation. In Asian Conference on Intelligent Information and Database Systems (pp. 519-528). Springer, Cham.
5. Sari, Y. A., & Adinugroho, S. (2017, August). Tomato ripeness clustering using 6-means algorithm based on v-channel otsu segmentation. In 2017 5th International Symposium on Computational and Business Intelligence (ISCBI) (pp. 32-36). IEEE.
6. Sari, Y. A., Adinugroho, S., Adikara, P. P., & Izzah, A. (2017, November). Multiplication of V and Cb color channel using Otsu thresholding for tomato maturity clustering. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 209-214). IEEE.
7. Sari, Y. A., & Adinugroho, S. (2018, May). Preprocessing of tomato images captured by smartphone cameras using color correction and V-channel Otsu segmentation for tomato maturity clustering. In 2018 5th International Conference on Electrical and Electronic Engineering (ICEEE) (pp. 399-403). IEEE.
8. Saravanan, G., Yamuna, G., & Nandhini, S. (2016, April). Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models. In 2016 International Conference on Communication and Signal Processing (ICCSP) (pp. 0462-0466). IEEE.

AUTHORS PROFILE



Yuita Arum Sari. Received her bachelor degree in Computer Science, University of Brawijaya (2011) and continued her Master degree in the Informatics Engineering in Institut Teknologi Sepuluh Nopember (2015). She also experienced of having mobility in Computer Science, Warsaw University of Technology (non-degree program in 2014-2015). Currently, she is a lecturer in Computer Science

in University of Brawijaya since 2016. She also an active member of Laboratory of Computational Intelligence and Computer Vision Research Group. Her research area is computer vision, image processing, machine learning, pattern recognition, and data mining.



Jaya Mahar Maligan. Received his bachelor degree in Faculty of Agriculture Engineering, University of Brawijaya in 2006 and continued his Master degree in the same institution in 2011. He is a lecture of Faculty of Engineering in University of Brawijaya since 2010. Her coverage research is about Nutrition metabolism. He is an active member in the Nutrition Laboratory in the same.



Sigit Adinugroho. Received his bachelor degree in Computer Science, University of Brawijaya (2012) and continued his Master degree in Computer Science, Uppsala University, Sweden (2015). Currently, he is a lecturer in Computer Science in University of Brawijaya since 2016. She also an active member of Laboratory of Computational Intelligence and Computer Vision Research Group. Her research area is computer vision, image processing, machine learning, pattern recognition, and data mining.



Yusuf Gladiensyah is an undergraduate student at the Computer Science Faculty, Brawijaya University, Indonesia since 2018. He received an award of champion of student in Innovation App Competition from University of Indonesia in 2019. His main research interest includes mobile application development, machine learning and image preprocessing, and archaeology based on mathematics. His current research is machine learning mobile application to identify nusantara ornament from candi and Indonesia ancient relic.