Estimating Food Volume in a Bowl Based on Geometric Image Features

by

Boyang Li

B.S. in Automation, Shannxi University of Science and Technology, 2018

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Boyang Li

It was defended on

March 24, 2021

and approved by

Mingui Sun, Ph.D, Professor

Department of Electrical and Computer Engineering

Zhi-Hong Mao, Ph.D, Professor

Department of Electrical and Computer Engineering

Ahmed Dallal, Ph.D, Assistant Professor

Department of Electrical and Computer Engineering

Liang Zhan, Ph.D, Assistant Professor

Department of Electrical and Computer Engineering

Thesis Advisor: Mingui Sun, Ph.D, Professor

Department of Electrical and Computer Engineering

Thesis Co-Advisor: Zhi-Hong Mao, Ph.D, Professor

Department of Electrical and Computer Engineering

Estimating Food Volume in a Bowl Based on Geometric Image Features

Boyang Li, M.S.

University of Pittsburgh, 2021

Image-based dietary assessment is important for health monitoring and obesity management because it can provides quantitative food intake information. In this thesis, a novel image processing method that estimates the volume of food within a circular bowl (i.e., the top rim of the bowl is a circle) is presented. In contrast to the Western culture where circular plates are most commonly used as food containers, circular bowls are the primary food containers in Asian and African culture. This thesis focuses on estimating the volume of amorphous food (i.e., food without a clear shape, such as a bowl of cereal) instead of food with usual shapes (e.g., an apple). Four geometric features of the food, namely food orientation, food area ratio, normalized curvature and normalized shape vertex, are extracted from 2D images. Based on these features, food volume is estimated using a linear or quadratic regression model. Our experiments show that, for 135 images of six different foods in a bowl of known shape, the mean absolute percentage error of our estimation was less than 20%, evaluated using a five-fold cross-validation technique.

Keywords: food volume estimation, Hough transform, image segmentation, curvature, regression.

Table of Contents

Preface . . . ix

1.0 Introduction . . . 1

1.1 Visual Data Collection . . . 3

1.2 Model-based Portion Size Estimation . . . 4

1.3 Thesis Outline . . . 6

2.0 Related Studies . . . 7

2.1 Image Segmentation . . . 7

2.1.1 Overview . . . 7

2.1.2 Algorithm Types . . . 8

2.2 General Equations of Ellipse . . . 11

2.3 Curvature . . . 12

2.4 Image Histogram . . . 14

2.5 Feature Extraction . . . 14

2.6 Hough Transform . . . 15

2.6.1 Basic Hough Transform . . . 15

2.6.2 Ellipse Hough Transform . . . 18

2.7 Machine Learning Regression Algorithm . . . 19

2.7.1 Overview . . . 19

2.7.2 Linear Regression Model . . . 19

2.7.3 Polynomial Regression Model . . . 20

2.8 Performance Metrics . . . 21

3.0 Proposed Methods . . . 23

3.1 Overview . . . 23

3.2 Data Collection . . . 24

3.3 Label . . . 24

3.4 Features . . . 25

3.4.1 Food Orientation 25
3.4.2 Food Area Ratio 26
3.4.3 Normalized Curvature 28
3.4.4 Normalized Shape Vertex 29
3.5 Regression 30
4.0 Results 32
4.1 Features 32
4.2 Regression 36
4.3 Discussion 37
5.0 Conclusions and Future Work 39
5.1 Conclusions 39
5.2 Future Work 39
Bibliography 41

List of Tables

1 Food Fullness. 25

List of Figures

1 Curvature definition, adopted from [1]. 13
2 Linear Hough transform. 16
3 Line Hough transform explanation, adapted from [2]. 17
4 Circle Hough transform, courtesy of MATLAB. 18
5 Examples of input images. 24
6 Food orientation. 26
7 Food area vs bowl area. 27
8 Food area ratio computing workflow. 28
9 Normalized curvature computing workflow. 29
10 Normalized shape vertex. 30
11 Normalized shape vertex computing workflow. 31
12 Feature results. 32
13 Food area ratio effects on fullness. 34
14 Normalized curvature/shape vertex effects on fullness. 35
15 Regression model test result. 36
16 Regression model five-fold validation result. 38

Preface

I would like to thank my advisor Dr. Mingui Sun and Dr. Zhi-Hong Mao for supervising me during my Master studies at the University of Pittsburgh. I would also like to express my gratitude to Dr. Wenyan Jia, Guangzong Chen, and Yiqiu Ren for their help throughout the research. Finally, I would like to thank my friends and family for their support during my two-year graduate study.

1.0 Introduction

There is an increasing concern about chronic health problems related to diet including obesity, hypertension, diabetes and heart disease. Diet education programs have been developed to advertise the effect of overweight and encourage people to have healthier dietary habits. Determining what someone eats during each meal provides valuable insights for addressing the dietary problem and helps to prevent many chronic diseases [3]. The need for a dietary assessment system to accurately measure dietary and nutrient intake becomes imperative. Also, evaluating precise dietary intake is considered as an open research problem in health and nutrition fields [4].

Traditionally, dietary assessment is usually conducted by self-reporting [5]. This type of dietary assessment is usually carried out in three methods: 24-hour dietary recall, food diary, and food frequency questionnaire.

The 24-hour dietary recall is a scientifically well-designed 20 to 60 minutes interview to record the total amount of each specific food and beverage consumption by the participants in the past 24 hours [6]. Instead of an automated self-assessment system, the interviews are typically administered by a trained interviewer. Interviewers will translate the description from the respondents into the energy and nutrients intake. The Foods Surveys Research Group (FSRG) of the United States Department of Agriculture (USDA) has made huge effort to make this method accurate [4]. The 24-hour dietary recall is precise because of the short interval, while the disadvantages of this method include high cost and long administration time.

Food diary is another well-known self-reporting method in which the research participant records his/her food intake as soon as the eating event completes [7]. The time interval of food diary usually lasts for a period of one or more days. Primary nutrients and food are recorded in a hand-written questionnaire, which will be evaluated and assessed by experienced experts to calculate the nutrients and food intake. Compared to the 24-hour dietary recall which is often used for population-based studies, food diary is a preferable dietary assessment method for clinical studies because it is generally more accurate [7].

Food frequency questionnaire (FFQ) is a dietary assessment tool that provides in a form of questionnaire to estimate the frequency of food item consumption and also the portion size information over a reference period [8]. Food frequency questionnaire has become a very important method for assessing dietary intake in epidemiological studies. Large health survey mainly gives a broadly representative roundup of the health condition within a specific population [9]. Compared with other dietary assessment methods, food frequency questionnaire is relatively economical and practical, easy and quick to operate [10].

As smartphones become more and more popular in the last 10 years. Mobile application could be used as a self-reporting method which helps to collect dietary supplement data over a shorter period of time. Respondents take pictures of each food before and after the meal (the after-meal picture is for possible left over). A number of mobile apps have been developed for dietary assessment using self-report. SapoFitness [11] is a conventional prototype of a mobile application for dietary evaluation. The application helps users to record daily diet and exercise. Also, SapoFitness could be used for intelligently setting diet plans based on customized objectives. Social network connection and alerting systems are included to help users to be more motivated during the process of controlling weight. Mobile applications could easily integrate many more functionalities to help people get healthier eating styles, but from a dietary assessment perspective, SapoFitness still uses a self-reporting conventional method.

Because of the complexity and diversity of different food recordings, self-reporting dietary assessment system is vulnerable to under-reporting. Portion size estimation is one of the main contributors to the under-reporting of the self-reporting dietary assessment process [4]. Respondents without enough training could lead to inaccurate results for the food volume estimation. Solid food portion size estimation could be significantly improved after training, while amorphous food volume is still hard to be estimated. Plus, self-report dietary assessment is time consuming and lack of consistency. Overall, although self-reporting provides the most intuitive and comprehensive way of dietary assessment data collection, automatic or semi-automatic monitoring systems are still indispensable.

One means of improving the data collected in these self-reporting processes is using technology. Technology assisted assessment could provide more consistent and high-quality results. Also, it could reduce the burden on both users and dietitians.

Computer vision based dietary assessment systems have been proposed and studied by a number of investigators in the last few years [12]. Two steps are usually involved into computer vision based dietary system workflow: visual data collection and portion size estimation.

Input data of computer vision based dietary systems could vary from images, video or 3D rangefinder. For image input, it could vary from a single image to multiple images from different angles. On the one hand, more visual input usually helps to improve the estimation result. On the other hand, it requires more human intervention which is not ideal for automatic dietary assessment system.

Portion size estimation is very hard because food shape and appearance could vary a lot due to food preparation conditions [3]. In majority, portion size estimation relies on 3D reconstruction model to transfer 2D image into 3D model with the correct scale. In these processes, a size reference (usually a checkerboard card) is placed next to the food to help estimate food scale from the image [4]. Model-based portion size estimation is a special case of portion size estimation, which evaluates the food volume based on the pre-defined and user-selected food 3D shape models.

1.1 Visual Data Collection

Visual data collection has been evolved in years and different methods and customized devices were proposed by different researchers. There are many existing technology-assisted diet assessment systems to proceed food images with a mobile device [13, 14]. DietCam [15] is a mobile system to help to evaluate food intake and aim to minimizing human intervention. A credit card needs to be first placed beside the plate, then users are required to take three pictures around the dishes approximately every 120 degree or shoot a short video. The 3D food model will be reconstructed based on the visual input. To compute the food size, a

credit card is used as the reference object since credit cards have a universally identical size. Users need to be involved to take pictures or shoot videos before and after meals. Since smartphones are the most popular and powerful mobile platform, the data collection process is appealing and not extremely hard. When the scene has less than six food items, DietCam could achieve a recognition accuracy of 92% which is quite promising. When the food shape and appearance is arbitrary (food residue), the predefined shape models are not feasible here.

Different dedicated wearable devices were proposed to help evaluate food, calories and nutrition intake [16, 17]. A wearable electronic device (eButton [18, 19]) is specifically used for objective dietary assessment data collection. eButton is a small electronic device containing a microphone, a miniature camera and several other sensors. It collects the visual data, which will be transferred in a predefined rate to the registered dietitian's computer for further image-based data processing. This device is designed to not affect users' life. This feature is particularly important as either questionnaire (self-reporting method) or mobile phone could tell if users are under an experiment for dietary assessment, which may lead to data inaccuracy.

A mobile structured light system (SLS [20]) contains a laser attachment and a mobile smartphone. Laser attachment projects grids on the food and the mobile phone will collect grid videos containing depth maps for the intersection points within each video frame. Depth map will be further combined with other 3D reconstruction techniques to reconstruct the 3D model of the food. Laser attachment is added in the system to provide more information for reconstruction from 2D image into 3D model. This system is burdensome and not suitable for daily use.

1.2 Model-based Portion Size Estimation

After food items are correctly segmented and identified, accurately estimating food volume in the image is the key step to determine the nutrient content of the food. Model-based portion size estimation is usually used for rigid or solid food without too much shape or appearance change. The basic idea is pre-defining or user-selecting a set of food items with

3D shape model templates, then trying to fit the 3D model into the object within the 2D image. The model template could be simple (e.g. sphere, cone) or complex (e.g. banana, pear).

Woo et al. proposed a method to automatically calculate portion size through volume estimation using a single image [14]. First, the camera calibration step computes camera intrinsic and extrinsic parameters, which would help to extract the geometry information in the 3D world. In this work, volume estimation method was developed for both spherical and prismatic shapes. With the assumption of the 3D model, feature points in 2D image are unprojected into the 3D world based on the parameters of the geometric class.

Jia et al. [21] used a wearable camera (eButton [19]) device to collect eating occasion information. It makes use of a known-size plate as the geometric reference. A few of simple geometric model shapes are predefined and manual adjustment is required.

3D/2D Model-to-Image Registration [5] is an improved method for estimating food volume from a single-view 2D image. The method utilizes the food global contour to resolve the position, scale and orientation of the user-selected and predefined 3D shape model. This method provides a robust solution for the simple food items, while doesn't give solutions to more complex 3D shape models. Also, this method only focuses on the outline of the food object and discards the internal structure of the food. (e.g. texture)

Food container mentioned in the previous work is mainly a circular plate, which is the most popular food container in Western countries. The developed dietary assessment system based on previous work theory would work well for Western country dietary research. In contrast, people usually use bowls instead of plates in Eastern countries. Model-based 3D reconstruction will be restricted as well since the side of the bowl will block the 3D shape model outline. Amorphous food is more common in Eastern cuisine. 3D predefined or user-selected shape models are inconsistent with the food object present in the image. This research focuses on estimating food volume within an circular bowl. (e.g. the top rim of the bowl is circular).

1.3 Thesis Outline

In Chapter 2, we review the related theories and specify the objectives of this work. It includes image segmentation, curvature, Hough transform, machine learning regression models and performance metrics, etc.

Chapter 3 details the methodology we opted for throughout the research. This chapter captures the details for data collection, features extraction and regression models we use in this research.

The results and discussion of this study are described in Chapter 4, emphasizing on the individual feature analysis, regression models and their performance evaluation. Also a detailed discussion related to the findings of this study based on the evaluation results using different regression models is included in this chapter.

Finally, we summarize this work and give future insights in Chapter 5.

2.0 Related Studies

2.1 Image Segmentation

2.1.1 Overview

An image contains lots of useful information. Thus, understanding the image or in other words extracting information from the image is an important technology. Image segmentation is an important topic in the field of computer vision. It is the first step in image analysis, the foundation of computer vision, and an important part of image semantic understanding. To be specific, image segmentation refers to the division of an image into several disjoint regions based on features such as grayscale, color, texture and geometric shapes, etc.. These features will be similar in the same region and show clear difference between different regions. In simple words, image segmentation is to separate the target from the background in an image.

The current image segmentation techniques include region-based segmentation, edge detection segmentation, segmentation based on clustering, and segmentation based on weakly-supervised learning in CNN (convolutional neural network), etc [2]. Edge detection segmentation and segmentation based on clustering are used in this thesis for data preprocessing.

In recent years, deep-learning based image segmentation has advanced really fast. The technology-related scene object segmentation, human body front background segmentation, face and human body parsing, three-dimensional reconstruction and other technologies have been widely used in many different industries such as security monitoring and self-driving systems.

2.1.2 Algorithm Types

There are many commonly used images segmentation algorithms, mainly under the scope of the following five categories: threshold segmentation method, regional growth segmentation, edge detection segmentation method, segmentation based on clustering and segmentation based on weakly-supervised learning in CNN, etc [2].

Threshold segmentation is a classic method in image segmentation. It uses the difference in pixel value between the target and the background from the grayscale image, and divides the pixel level into several categories. Grayscale image will be converted into a binary image with the threshold enforcement. Threshold segmentation can usually be divided into two categories: local threshold method and global threshold method. The local threshold method needs to select multiple segmentation thresholds and divides the image into multiple target regions and background regions [2]. The global threshold method divides the image into two regions of target and background by a simple threshold [22].

The most critical step of this algorithm is to solve the optimal gray threshold according to a certain criterion function. The Ostu algorithm [23] is considered to be the most commonly used and the best algorithm for threshold selection in image segmentation. It is simple to calculate and is not affected by image brightness and contrast. Another outstanding threshold segmentation method is KSW entropy algorithm. Entropy represents the amount of information within the image. The greater the amount of image information is, the larger the entropy will be. Besides these two methods, there are fuzzy set method, and moment preserving method, etc.

The threshold segmentation method is simply to compute and can always use closed and connected boundaries to define non-overlapping areas. Generally speaking, images with stronger contrast between background and the target can get a better segmentation result.

The basic idea of regional growth segmentation is to combine pixels with similar attributes to form regions. Region-based segmentation is a technique to partition an image into regions directly, where two steps are usually involved: to determine the initial seed points first, and then merge the surrounding neighborhood with similar attributes to the pixels in the region.

Edge detection segmentation focuses on the measuring, monitoring and positioning of the gray level changes of the image. The method aims to extract the features from the discontinuous parts in the image. Algorithms for edge detection usually contain four steps: filtering, enhancement, detection, and localization. There are some first derivative edge detectors such as Roberts Operator, where G_x and G_y are calculated using the following kernels

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{2-1}$$

The Roberts cross operator provides a simple approximation to the gradient magnitude [24]. Sobel Operator is another commonly used edge detection operator. It places an emphasis on pixels that are closer to the center of the mask. Then S_x and S_y can be calculated using convolution kernels

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{2-2}$$

The Prewitt operator uses the same equations as the Sobel operator does, except that the constant $c = 1$. This operator does not place any emphasis on pixels that are closer to the center of the kernels, therefore [24]

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}. \tag{2-3}$$

Comparing these three edge detection operators, the effect of the Sobel operator is the best, because the influence of the Sobel operator on the position of the pixel is weighted.

Laplacian operator is an isotropic second order differential operator. It is more appropriate to be used when the main concern is to find the position of the edge regardless of the pixel gray scale difference around it [25]. In addition, Laplacian operator could not be used for finding the orientation of edge. In the presence of noise, the Laplacian operator needs to be accompanied with a low-pass filter before detecting the edge [2]. Therefore, the usual

segmentation algorithm combining with the Laplacian operator and the smoothing operator is used to generate the new template [2]. The formula of the Laplacian of a function $f(x, y)$ is

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{2-4}$$

Moreover, the Laplacian operator can be expressed as following

$$\nabla^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2-5}$$

Sometimes, it is better to give more weight to the center pixels in the neighborhood [25]. Therefore, the Laplacian operator can be expressed as

$$\nabla^2 = \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}. \tag{2-6}$$

The Laplacian operator is used to improve the blurring effect since it conforms to the descent model [2].

The method of merging similar pixels in grayscale images and color images is called clustering. The image is represented as different regions by clustering, which is the so-called clustering segmentation methods. This method is to partition the feature space according to their aggregation in the feature space, and then map them back to the original image space to obtain the segmentation result.

K-means is one of the most commonly used clustering algorithms, which uses distance as a similarity evaluation index. The basic idea of the algorithm is to put samples into different clusters according to distance. The closer the distance between two points gets, the greater the similarity will be. As a result, compact and independent clusters can be obtained as the clustering target [2]. The implementation process of K-means is expressed as follow [2]

- Initialize k constant and random-selected initial cluster centers.

- Calculate the distances from each sample to all cluster center and then classify each sample to the nearest cluster center.
- For each cluster, take the mean of all samples as the new center of the cluster.
- Repeat steps 2 to 3 until the cluster center no longer changes or the preset number of iterations is reached, then output the final cluster center and the category to which each sample belongs.

The biggest advantage of this algorithm is its simplicity and fast calculation speed, and the key to the algorithm is the selection of the initial center and the distance formula.

Deep learning has achieved many breakthroughs and been widely used in many fields such as image classification, detection and segmentation. Lin et al. proposed a weakly supervised learning method based on Scribble marking [26]. ScribbleSup contains two steps. The first step is to spread the pixel category information from scribbles to other unlabeled pixels and automatically completes the labeling of all training images; the second step is to use these labeled images to train CNN. Papandreou et al. has further studied the use of bounding box and image-level labels as labeled training data based on DeepLab [27]. The Expectation Maximization Algorithm (EM) is used to estimate the category of unlabeled pixels and the parameters of the CNN [2].

2.2 General Equations of Ellipse

The standard equation for an ellipse is as follow:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{2-7}$$

It represents an ellipse centered on the origin with width $2a$ and height $2b$. In most cases, the ellipse is centered at a random point, and its axis is not parallel to the coordinate axis. In this scenario, it is always possible to obtain the ellipse function from an ellipse in a standard position and then implement rotation or translation. Therefore, by applying rotation and translation to the standard equation of an ellipse the equation of any ellipse can be found.

Using the equations to rotate a point about the origin by angle α clockwise the inverse operation can be obtained by rotating through $2\pi - \alpha$. Then combining the standard ellipse equation the rotated ellipse equation can be expressed as

$$\frac{(x\cos\alpha + y\sin\alpha)^2}{a^2} + \frac{(x\sin\alpha - y\cos\alpha)^2}{b^2} = 1. \tag{2-8}$$

2.3 Curvature

Curvature is a geometry quantity to describe the degree of curvature such as the degree to which curved surface deviates from a plane or the degree to which a curve deviates from a straight line.

A plane curve is defined by the equation $y = f(x)$ and and the tangent line is drawn to the curve at point $M(x, y)$. The tangent line then forms an angle α with the horizontal axis.

Figure 1 [1] shows that, there are two tangent lines at point M and M_1 respectively [1]. When point M moves to M_1 along the curve, the distance traversed is Δs and the angle between the corresponding tangent line and the horizontal axis changes from α to $\alpha + \Delta\alpha$. Therefore, as the point moves by distance Δs, the tangent rotates by the angle $\Delta\alpha$ [1]. We can then obtain the curvature of the curve at the point M

$$K = \lim_{\Delta s \to 0} \left| \frac{\Delta\alpha}{\Delta s} \right|. \tag{2-9}$$

According to the Equation 2-9, the curvature at a point of the curve represents the rotation speed of the tangent point of the curve at that point.

If a curve is defined as an equation $y = f(x)$, the curvature at a point $M(x, y)$ can then be expressed in terms of the first and second derivatives of the function $f(x)$ and has the following form [1]

$$K = \frac{|y''(x)|}{[1 + (y'(x))^2]^{\frac{3}{2}}}. \tag{2-10}$$

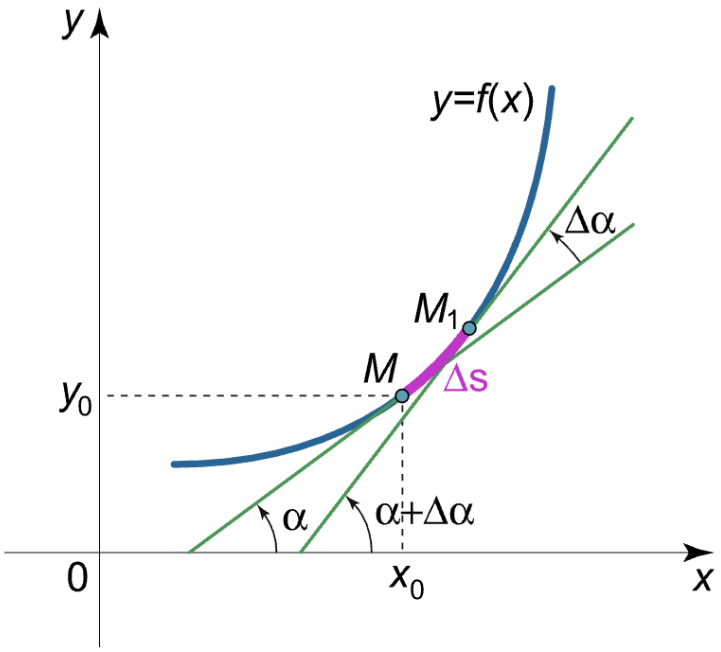


Figure 1: Curvature definition, adopted from [1].

If a curve is defined as the polar equation $r = r(\theta)$, the curvature at the point $M(x, y)$ is expressed as [1]

$$K = \frac{|r^2 + 2(r')^2 - rr''|}{[r^2 + (r')^2]^{\frac{3}{2}}}, \tag{2-11}$$

where the curvature is the reciprocal of radius of curvature, which means the curvature is

$$K = \frac{1}{R}, \tag{2-12}$$

where R is the radius of curvature. Thus, the radius of curvature at a point $M(x, y)$ can be expressed as the following form [1]

$$R = \frac{[1 + (y'(x))^2]^{\frac{3}{2}}}{|y''(x)|}. \tag{2-13}$$

2.4 Image Histogram

An image histogram is a statistical table that reflects the distribution of pixels in an image. The horizontal axis of the table represents the pixel value of the image, which can be grayscale image or colored image. The vertical axis represents the total number of pixels of each pixel value in the image or the percentage of all pixels. Since the image is composed of pixels, the histogram reflecting the pixel distribution can often be used as a very important feature of the image.

2.5 Feature Extraction

Feature extraction is a widely used computer vision and image processing technic to extract non-image descriptions from image data. Usually, the input data of the algorithm is very large and often cannot be directly processed. Thus, it can be converted into non-redundant features vector. This process of converting original input data into feature sets is

called feature extraction. In pattern recognition and image processing, feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space [28].

In general, each image has features that can be distinguished from other images. Examples of these features include those that can be obtained directly from the image itself, such as brightness and texture, and those that need to be obtained through certain transformations such as histograms, etc. The selected features must not only describe the image well, but more importantly, it should be able to distinguish different types of images well.

Feature extraction is an important step in the construction of any pattern classification and aims at the extraction of the relevant information that characterizes each class [28].

2.6 Hough Transform

2.6.1 Basic Hough Transform

Hough transform is a feature-extraction technology in image processing, which uses voting algorithm to detect objects with specific shapes. This process is carried out in a parameter space. A set conforming to the specific shape is obtained by calculating the local maximum of the cumulative result, and it is then taken as the Hough transform result. The Hough transform was first proposed by Paul Hough in 1962 [29] and later popularized by Duda and Hart in 1972, who called it a "generalized Hough transform" [30]. At first, the classical Hough transform was used to detect straight lines within the image. Later it was extended to detect the objects of arbitrary shapes, mostly circles and ellipses.

The linear Hough transform algorithm estimates the two parameters that define a straight line. For example, if (ρ, θ) are used to represent a straight line, then ρ is the distance from the line to the origin and θ is the angle between the orthogonal line and the x-axis. Figure 2 shows this expression.

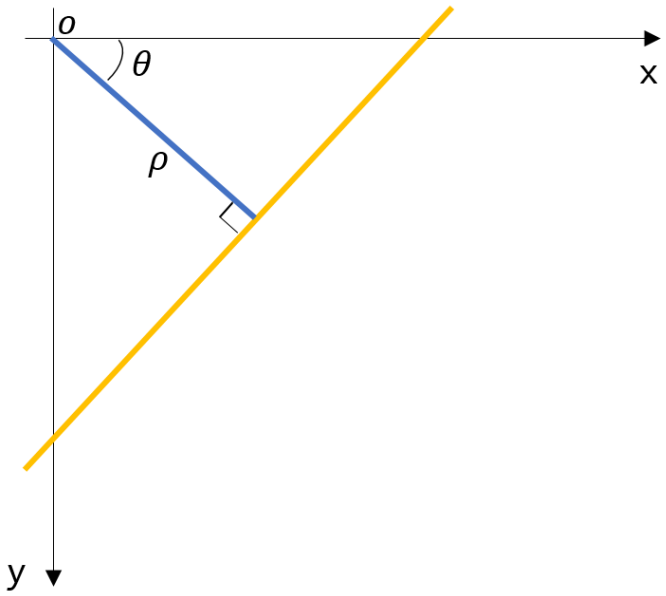
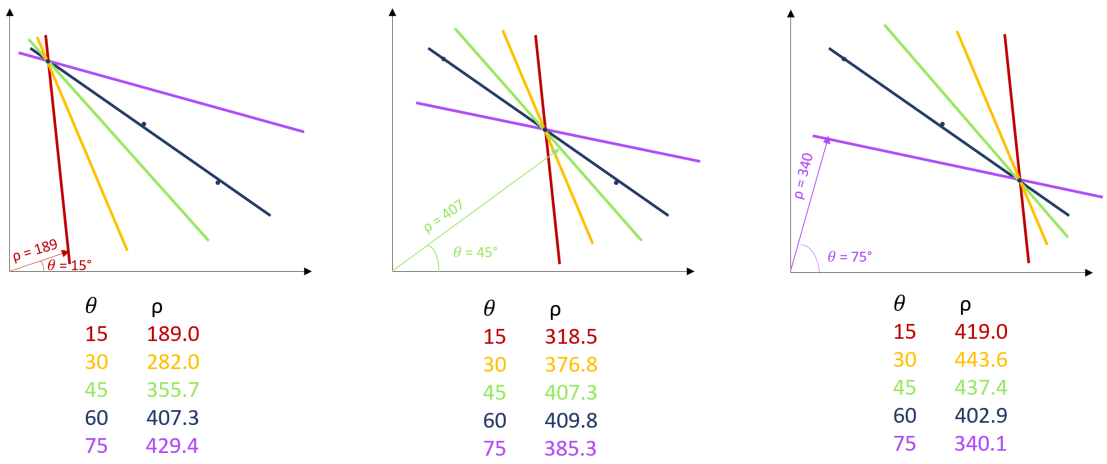


Figure 2: Linear Hough transform.

Figure 3 shows the Linear Hough transform diagrams. The main idea of using Hough transform to detect straight lines is: suppose a straight line in n directions for each point, and calculate the (ρ, θ) coordinate of the n straight lines to obtain n coordinate points. If there are N points to be judged, the number of (ρ, θ) coordinate will be $(N \times n)$. Regarding to these $(N \times n)$ number of (ρ, θ) coordinates, θ is a discrete angle with a total of 180 values. If multiple points are on a straight line, then the θ of these points must be equal and the ρ of these points are also approximately equal.

There are three points shown in the Figure 3. And for each data point, a number of lines are plotted going through it, all at different angles. The perpendicular distance of each line from the origin and angle of each support line are calculated. From the result, it can be seen that in either case the support line at $\theta = 60°$ has similar r. It's clear that the support lines which shown in blue ones are very similar, which illustrates that all three data points lie close to the blue line.



θ	ρ
15	189.0
30	282.0
45	355.7
60	407.3
75	429.4

θ	ρ
15	318.5
30	376.8
45	407.3
60	409.8
75	385.3

θ	ρ
15	419.0
30	443.6
45	437.4
60	402.9
75	340.1

Figure 3: Line Hough transform explanation, adapted from [2].

The use of the Hough transform to detect circles was first outlined by Duda and Hart [30]. If the circle is parameterized by its center coordinates (a, b) and radius r, then these three parameters correlate to the position of edge points (x, y) and form a circle via a constraint [31]

$$(x - a)^2 + (y - b)^2 = r^2. \tag{2-14}$$

Figure 4 shows a circle Hough transform detection. When detecting a circle with a certain radius, a circle in the circle image space corresponds to a point in the Hough parameter space and a point in the Hough parameter space corresponds to a circle in the image space. If the parameters of the points on the same circle in the circle image space are the same, then their corresponding circles in the Hough parameter space will pass the same point. According to the degree of aggregation of the points in the Hough parameter space, after transforming all the points in the original image space to the parameter space, it'll be clear whether or not there is a figure similar to a circle in the image space.

When the radius of the circle is unknown, it can be regarded as the detection of a circle with two parameters - the center and the radius. Each point in the image space corresponds to a cluster of circular curves in the Hough parameter space, which is actually a cone.

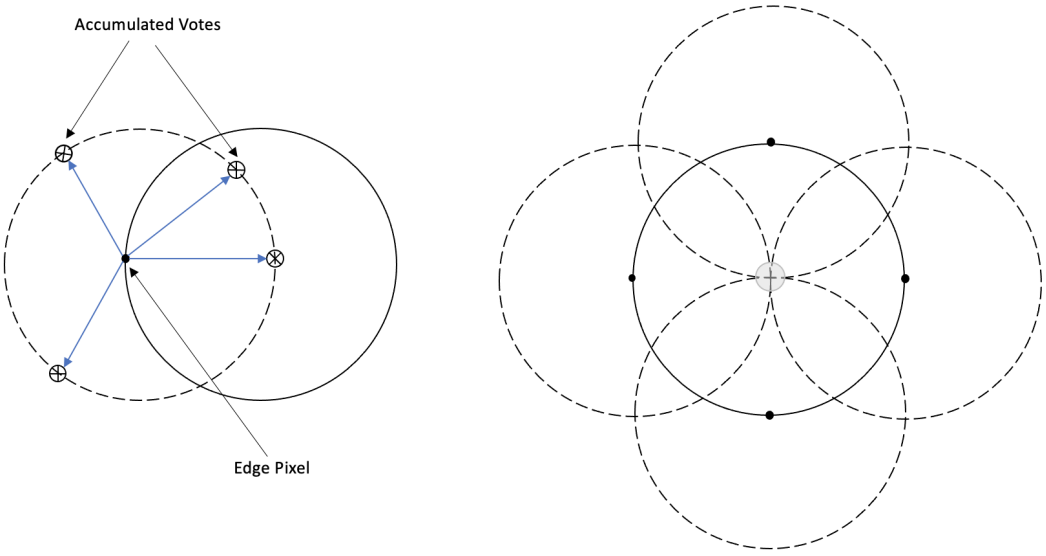


Figure 4: Circle Hough transform, courtesy of MATLAB.

2.6.2 Ellipse Hough Transform

By improving the Hough transform algorithm, it can be used to detect ellipses too. In this case, there are four parameters need to be known, which are the center of the ellipse, orientation, major and minor axis. Therefore, the dimension of the parameter space will increase to 4 and the amount of calculation will also be increased.

Our group previous research proposed a new improvement method to detect the ellipse information in images. The steps are as follows:

- Rotate and stretch the image, in order to use Circle Hough transform to find ellipse as circle.
- After finding the circle, locate the ellipse in original image by calculation.

2.7 Machine Learning Regression Algorithm

2.7.1 Overview

Regression analysis is a method of predictive modeling technology that studies the relationship between the outcome variable and the predictors. This technique is used in forecasting time series models and finding causal relationship between variables. In the same set of data predicted in regression analysis, there is a certain relationship between different variables that can be quantified. To obtain the functional expression of this relationship, we use the statistical techniques and it's believed that this relationship can be deduced from the sample if there are enough data samples. Due to the uncertainty of this reverse deduction, we need to make multiple assumptions and then verify it. The most important feature in regression analysis is that the predicted results are continuous.

2.7.2 Linear Regression Model

Linear regression is the most well-known and simplest modeling technique among all the regression models. In a linear regression model, the dependent variable is continuous and the independent variable can be continuous or discrete. Linear regression establishes the relationship between the dependent variable Y and one or more independent variables X by using the best fitting straight line, which is also called the regression line.

In the general linear regression model, the random variable y_i satisfies

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \qquad i = 1, \ldots, n, \tag{2-15}$$

where ${x_i}^T\beta$ is the inner product between vectors x_i and β. The n equations together with the assumptions about ϵ_i can be written in matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim (0, \sigma^2 \boldsymbol{I}_n), \tag{2-16}$$

where $\boldsymbol{y}$ is a vector of observed values $y_i (i = 1, \ldots, n)$ of the variable called the regressand. $\boldsymbol{X}$ can be seen as a matrix of the independent variables. $\boldsymbol{\beta}$ is a vector of regression parameters and $\boldsymbol{\epsilon}$ refers to a vector of independent and identically distributed noise.

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \quad \boldsymbol{X} = \begin{pmatrix} x_1^T \\ x_2^T \\ \cdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_n \end{pmatrix} \tag{2-17}$$

The advantages of linear regression models are:

- Modeling is fast and simple, which is especially suitable for situations where the relationship to be modeled is not very complex and the amount of data is small.
- Linear regression model is very sensitive to outliers.
- It is easy to give a more intuitive understanding and explanation

2.7.3 Polynomial Regression Model

Although the linear regression model is the simplest model, it has many assumptions. One of the most important one is to assume a linear relationship between the response variable and the explanatory variable. However, in many real datasets, the linear relationship is often weak and even non-exist. Therefore, some nonlinear regression models have been proposed. Polynomial regression is a linear combination polynomial that converts a feature into a high-order feature. It is thus a curve that fits into the data points.

Polynomial regression is a special case of multiple regression with only one independent variable X. One-variable polynomial regression model can be expressed as the following[32]

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_k x_i^k + e_i \qquad i = 1, \ldots, n, \tag{2-18}$$

where k is the degree of the polynomial, which is also called the order of the model.

2.8 Performance Metrics

The regression task is to use other relevant independent variables to predict the state of the outcome variable at a specific point. Unlike classification tasks, regression tasks output continuous values within a given range. According to the problem to be solved, researchers use different performance indicators such as mean squared error MSE, mean absolute percentage error $MAPE$, R-squared R^2 (coefficient of determination), etc.

- Mean squared error MSE:

$$MSE = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2. \tag{2-19}$$

Mean squared error is the most used performance indicator of the regression system. It is the average of the squared difference between the target value and the value predicted by the regression model. In the above equation, y_i is the actual value and $\hat{y}_i$ is the predict value.

- Mean absolute percentage error $MAPE$:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|. \tag{2-20}$$

Mean absolute percentage error $MAPE$ can measure relative performance, which is a useful measure to compare the accuracy of prediction between different items. This metric is the expected value of the relative error loss.

- R-squared R^2:

$$R^2\ Score = 1 - \frac{\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1}(y_i - \bar{y}_i)^2}. \tag{2-21}$$

R^2 is also called the coefficient of determination, which reflects the degree of interpretation of the independent variable to the change of the dependent variable. The R^2 score is always between zero and one. The closer it is to 1, the better the model fits. When the R^2 value is 0.5 or below, the regression explains only 50% or less of the variation in the data and the prediction may be poor [32].

- Explained variance score:

$$Explained\ Variance\ Score = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}}. \tag{2-22}$$

Explained variance score indicates how close the dispersion of the difference between all predicted values and the sample is to the dispersion of the sample itself.

3.0 Proposed Methods

3.1 Overview

In this research, we use a machine learning process to solve the food volume estimation problem within a circular bowl. Common machine learning steps consist of data collection, data preparation, model selection, model training, model evaluation, parameter training and prediction steps. Data is collected with a mobile phone with known camera parameters. Camera calibration is used to estimate the parameters of a lens or video cameras. These parameters will then be used to correct lens distortion, which helps to measure the size of an object in world units correctly.

Instead of using every individual pixel information within the captured food image, features are extracted based on the overall pixel information in the bowl. For the first step, we segment pixels within the bowl with improved Hough transform and food pixels with k-means image segmentation algorithm. This work concentrates on examining volume estimation and assumes that food items have been properly segmented. One manual segmentation method is added to ensure food pixels are segmented properly.

Afterwards, bowl geometric features and food volume related features are extracted from the segmented pixel information. We define four different features in our work to represent food volume: food orientation, food area ratio, normalized curvature, and normalized shape vertex. Food orientation is a geometric feature which demonstrates the photo shooting angle and position. Food area ratio demonstrates the ratio of food pixels to the number of pixels within the bowl, which shows the food amount to a certain extent. Normalized curvature and normalized shape vertex demonstrate the bulge of the amorphous food outline contour because amorphous food will present a "hilly" shape as we add more food into the bowl.

In this work, machine learning training process regression model is used instead of classification model. Fundamentally, classification is used for predicting a label and regression is about predicting a quantity. Regression models could give us more insight and more accurate food volume within the bowl. The definition of volume we use in this work is the same as the

Figure 5: Examples of input images.

volume definition in Physics: three-dimensional space occupied by a substance or enclosed by a surface. Food volume is solved with density and mass. To be more specific, we adopt both linear regression model and polynomial regression model, which require the least amount of computing resources. These two regression models will provide us more flexibility if we want to integrate the machine learning process into a real-time dietary assessment system.

3.2 Data Collection

We use 6 different kinds of Chinese food in our dataset with 135 images in total. Each food has 4 different volumes. From Figure 5, we can see the food we used in Chinese cuisine is usually amorphous and without a clear shape.

3.3 Label

From the Table 1, we can see that the weight of the empty bowl is 340 grams and the volume is 620 ml. The right image within Figure 5 demonstrates how we collected the weight

Food Name (#)	Food Density (g/ml)	Bowl Weight (g)	Total Weight (g)	Net Food Weight (g)	Food Volume (ml)	Bowl Volume (ml)	Fullness
#1	0.57	340	654	314	547	620	0.88
#1	0.57	340	567	227	395	620	0.64
#1	0.57	340	485	145	252	620	0.41
#1	0.57	340	388	48	84	620	0.14
#2	0.91	340	796	456	501	620	0.81
#2	0.91	340	698	358	393	620	0.63
#2	0.91	340	601	261	287	620	0.46
#2	0.91	340	490	150	165	620	0.27

Table 1: Food Fullness.

of the food for a certain image. According to the weight and density of the food we measured, we can get the ratio of the volume of the food to the total volume of the bowl. We define the ratio as Fullness, which will be treated as the regression model label during the machine learning step.

$$Fullness = \frac{V_{Food}}{V_{Bowl}} \tag{3-1}$$

3.4 Features

3.4.1 Food Orientation

Food orientation is a key feature to demonstrate the distortion of the bowl because of the different photo shooting angles. This distortion also applies to food outline contours. Since amorphous food outline contours are usually irregularly shaped, estimating distortion based on food outline contours is very challenging. On the other hand, since the rim of a bowl is usually round we could use bowl contour as a reference to estimate the distortion of the image.

Within the image segmentation step, we use improved and adaptive Hough transform to find the bowl contour. The adaptive Hough transform mentioned in the related theory part could handle the ellipse detection case within a 2D image. The angle between optic axis

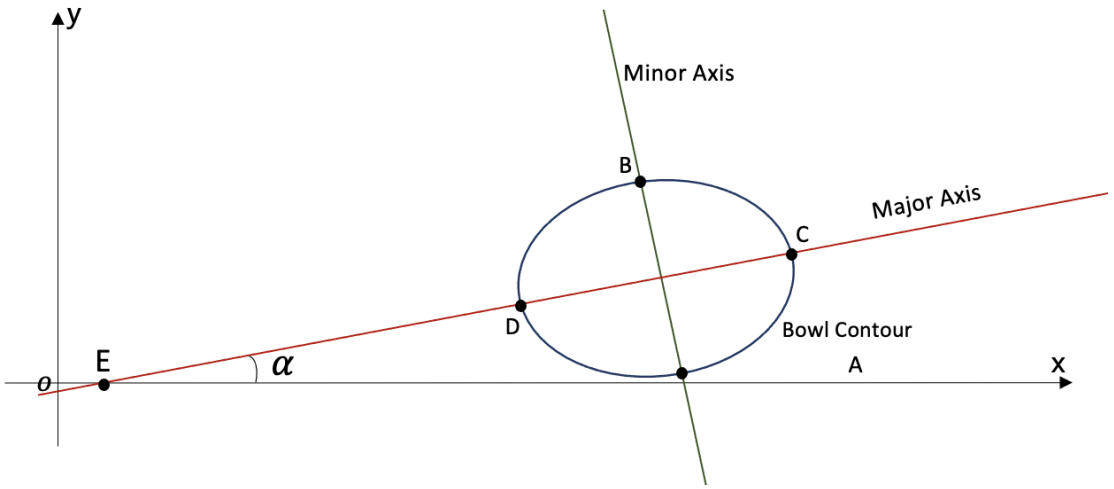


Figure 6: Food orientation.

and vertical direction could significantly affect the ellipse shape. Also, the distance between camera and the food is another contributor. Reversely, we also want to develop a feature to effectively demonstrate the camera position and the shooting angle. Food orientation is defined as the ratio of the ellipse major axis length to the ellipse minor axis length, which refers to the ratio of L_{CD} to L_{AB} in Figure 6. Larger food orientation means a flatter ellipse, while smaller food orientation means a more circular ellipse.

$$Food\ Orientation = \frac{L_{Major\ Axis}}{L_{Minor\ Axis}} \tag{3-2}$$

3.4.2 Food Area Ratio

Food area ratio is another key feature to estimate the volume of the food in a bowl. The larger the area occupied by food in the 2D image, the larger the amount of food there is. Since different camera shooting distances and angles will significantly affect the amount of the food pixels, we use the ratio between the number of pixels occupied by the food part and the total pixels within the bowl as the food area ratio. This feature is used to approximately estimate the amount of food in the bowl. This feature can be expressed as the following

Figure 7: Food area vs bowl area.

form

$$Food\ Area\ Ratio = \frac{N_{Food\ Pixels}}{N_{Total\ Pixels}}. \tag{3-3}$$

Figure 7 shows that the pixels surrounded by green line refer to food pixels, while the pixels surrounded by red line refer to bowl pixels. Food area ratio refers to the ratio of food pixels to bowl pixels.

Figure 8 shows the workflow we used to compute the food area ratio feature. For the first step, we obtained the number of pixels within the bowl by performing adaptive Hough transform on the 2D image. Then, food pixels were extracted with k-means image segmentation method. Since this work concentrates on examining volume estimation and assumes that food items have been properly segmented, manual image segmentation with GIMP is used to obtain the ideal food segmentation result when k-means couldn't. At last, food area ratio is calculated based on Equation 3-3 for each 2D image.

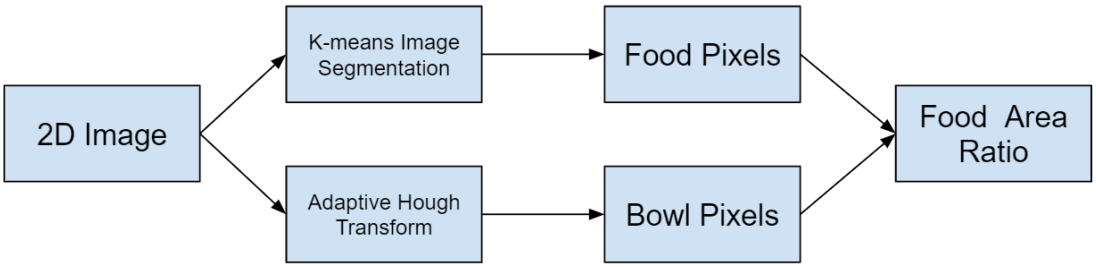


Figure 8: Food area ratio computing workflow.

3.4.3 Normalized Curvature

Usually, when we add more food into the bowl, the amorphous food will apply a "hilly" shape. The bulge of the food within the bowl affects the accurate estimation of the food volume. Therefore, we want to find an appropriate parameter to describe the bulge of the food. In this case, food contour outline, especially upper contour outline, is extracted to evaluate the bulge of the food. Curvature demonstrates the amount by which a curve deviates from a straight line, and curvature is a good measurement to describe how curvy a line or plane is. Based on this, the maximum curvature for the convex upper contour is first selected as one of the features to estimate food volume.

According to the concept of the curvature, the scale of maximum curvature could vary a lot which is against the stability of the classification or regression model in later steps. We came up with a meaningful way to normalize the maximum curvature of the food upper contour. Normalized curvature is proposed as the product of maximum curvature of the food upper contour and square root of the food area in 2D image, which is shown in Equation 3-4.

$$Normalized\ Curvature = Maximum\ Curvature \times \sqrt{Number\ of\ Food\ Pixels} \quad (3\text{-}4)$$

Figure 9 demonstrates the workflow for computing normalized curvature. K-means image segmentation method is first performed to segment food areas out of the background. Since we mainly focus on the outline of the food contour, Canny edge detection is used to get

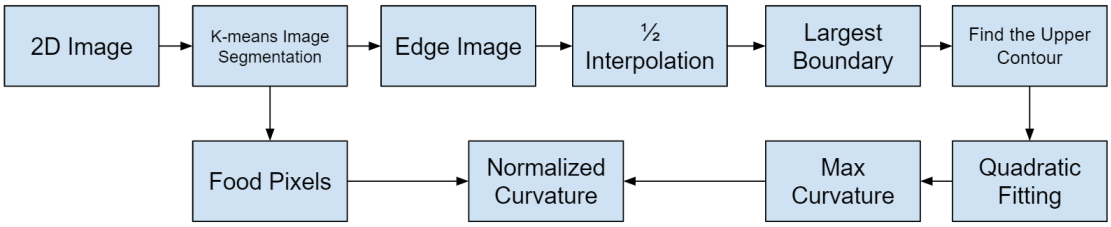


Figure 9: Normalized curvature computing workflow.

the edge information of the binary segmentation image. Either because segmentation result is not ideal or there is small food residue on the side of the bowl, we usually find multiple closed areas detected by the food segmentation step. We usually refer food area to the largest closed area in the segmentation image.

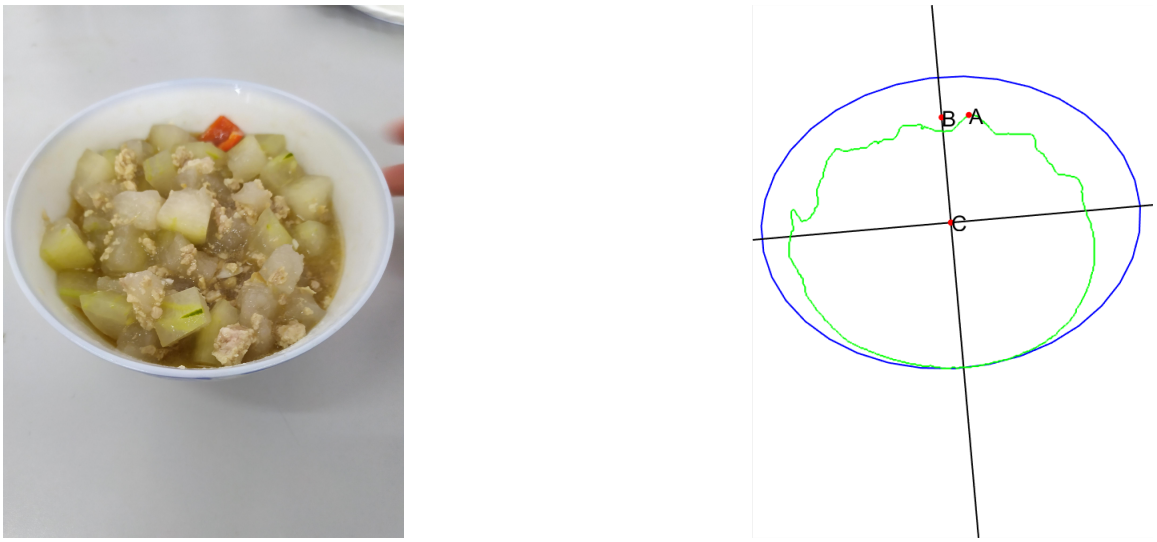
To find the largest closed boundary, we first perform $\frac{1}{2}$ scale interpolation to make every boundary point continuous. After this, we set a horizontal line to split the food area into two parts so that we can get the upper contour of the food area. After obtaining the boundary of the upper half of the food contour, we use the quadratic curve fitting method to fit the boundary. Finally, we can find the maximum curvature based on the fitted curve function.

Normalized curvature is designed to be a feature evaluating the bulge of the food. If the containing food is liquid, the outline of the food is usually either a circle or an ellipse because of distortion. In circular cases, normalized curvature is constant since circle radius is inversely proportional to the curvature. Based on this, if the "hilly" shape of the food is sharper, the normalized curvature will be larger.

3.4.4 Normalized Shape Vertex

Normalized shape vertex is the last extracted feature which is used to calculate the food volume. When the food volume is greater than the bowl volume, which is possible for amorphous food, normalized shape vertex is also a parameter to demonstrate the bulge of the food upper contour. Figure 10b shows the definition of the normalized shape vertex. Point A refers to the highest point of the food upper contour and B is the projection point

of A onto the minor axis of the bowl ellipse. Point C refers to the center of the ellipse. Normalized shape vertex is equal with the distance between B and C over the length of minor axis of the bowl ellipse.



(a) Input image.

(b) Definition.

Figure 10: Normalized shape vertex.

Figure 11 shows the workflow for computing the normalized shape vertex. We can get the food contour outline with the same method mentioned in the normalized curvature section. Minor axis line function could be computed based on the ellipse parameters (center coordinates, rotation angle, etc.) generated by adaptive Hough transform. With this information, we could project all the points within the food contour outline onto the ellipse minor axis, and then obtain the highest projection point.

3.5 Regression

Regression model is adopted over classification model since we aim to obtain more accurate food volume estimation. Linear regression model is first chosen because of the features

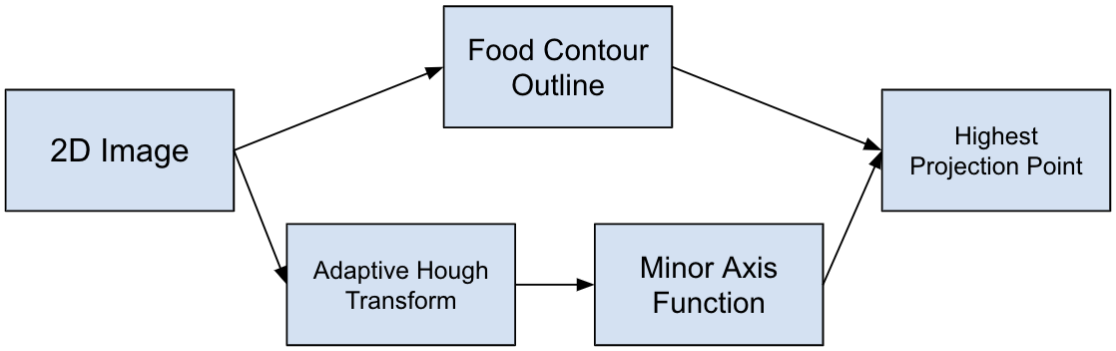


Figure 11: Normalized shape vertex computing workflow.

described above, especially food area ratio and normalized shape vertex, share the same trend as the food volume. Therefore, we first use the multivariate linear regression model which is described in Equation 3-5, where y refers to dependent variable (Fullness), and x_i refers to explanatory variables (features extracted from 2D images). β_0 is the y-interception constant term and β_p is the slope coefficients for each explanatory variable. ϵ refers to the linear regression model's error term (residue), where our objective is to minimize the residue.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \tag{3-5}$$

In addition to the multivariate linear regression model, we also used the quadratic regression model to analyze the data. As the workflow demonstrated in previous sections, different features may be correlated with each other. Therefore we take all the cross terms into account while transforming a quadratic regression problem into a multivariate linear regression model. 2-variable quadratic regression model could be described as Equation 3-6 and we will extend this into 4-variable quadratic regression model in this work.

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2 + \epsilon \tag{3-6}$$

Mean squared error, mean absolute percentage error, R-squared, and the explained variance score are the error metrics we use to evaluate the performance of the trained regression models.

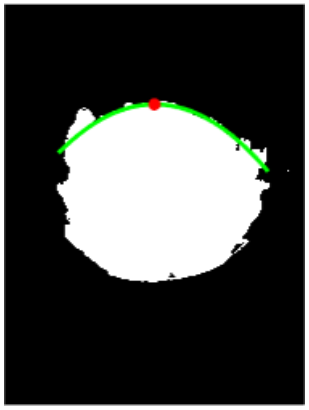
4.0 Results

4.1 Features

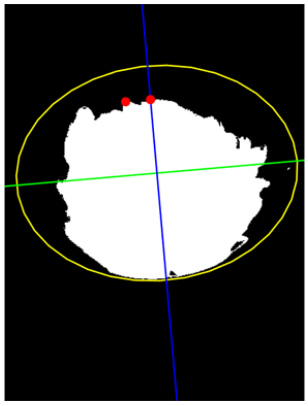
(a) Input image.

(b) Bowl detection.

(c) Food segmentation.



(d) Max curvature.



(e) Peak projection.

Figure 12: Feature results.

Fig 12 shows the features extracted from a single 2D food image. Adaptive Hough transform is used to obtain the ellipse shape of the bowl contour, shown in Fig 12b. With the parameters of the ellipse, we could calculate the orientation as the ratio of the length of the ellipse major axis and the length of the ellipse minor axis.

Fig 12c demonstrates the food segmentation result based on k-means image segmentation method, which helps us get the amount of food pixels within the image. Food area ratio feature is calculated as the ratio of the number of food pixels over the number of pixels within the bowl.

Based on the food segmentation result, we could obtain the food contour and perform quadratic fitting on the food upper contour to get the maximum curvature point (the red point shown in Fig 12d). Then we could calculate the normalized curvature value using the product of the maximum curvature and the square root of the number of the food pixels.

Fig 12e shows the shape vertex projection computing result, where the shape vertex point on the food contour is projected onto the minor axis of the ellipse. Then normalized shape vertex is computed using the ratio of the distance between the shape vertex projection point and the ellipse center over the ellipse semi-minor axis length.

Food area ratio is the most intrinsic feature to represent the food volume. Usually, when the food area ratio is larger, the Fullness is larger. Fig 13 shows the effects of the food area ratio on the Fullness for a certain food (Stir fry Pork with Tofu skin). As we can see in the figure, food area ratio and Fullness usually share the same trend.

Fig 14 shows the maximum curvature and the shape vertex projection results for the same food (Stir Fry Pork with Cowpea) with different volumes. Maximum curvature is computed based on the food upper contour outline. When the shot angle is flat, the food upper contour outline demonstrates the bulge of the hilly shape of the food, which is shown in the second rows of Fig.3. When the shot angle is high, the food upper contour outline shows the arc shape of the boundary between food and the side of the bowl. That's also the reason why we added the food area factor into this feature to demonstrate the food volume while the shot angle is high.

Similarly, normalized shape vertex is extracted to show the bulge of the food while the shot angle is flat, which is shown in the third row of Fig 14. When the shot angle is high, normalized shape vertex is used to demonstrate the food volume directly.

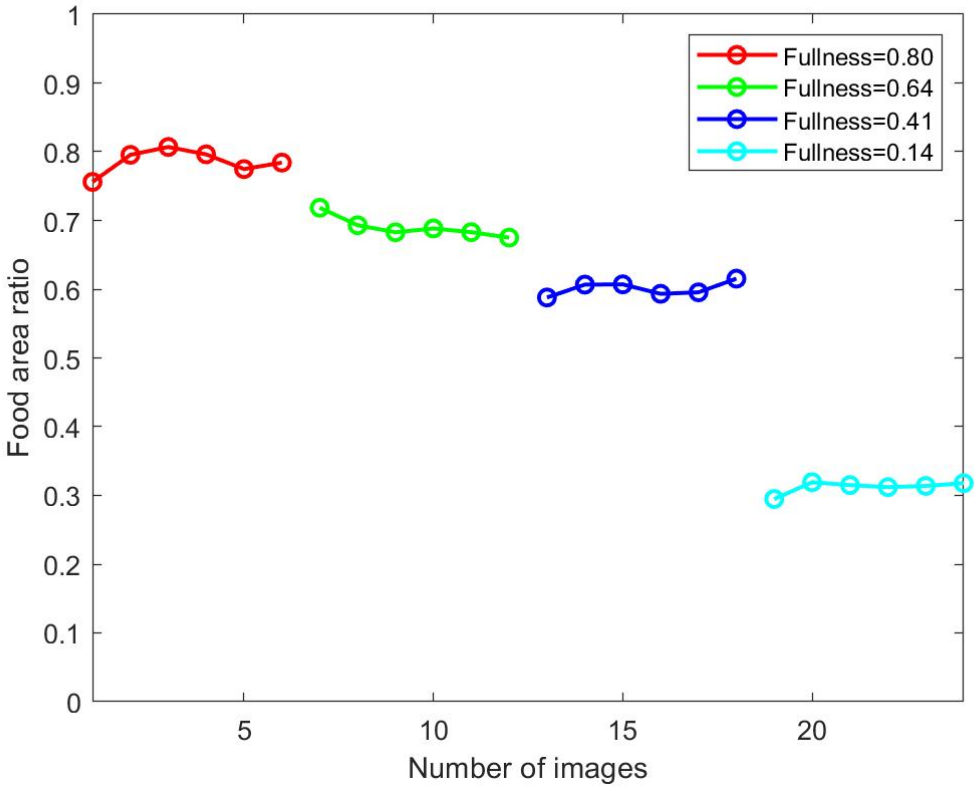


Figure 13: Food area ratio effects on fullness.

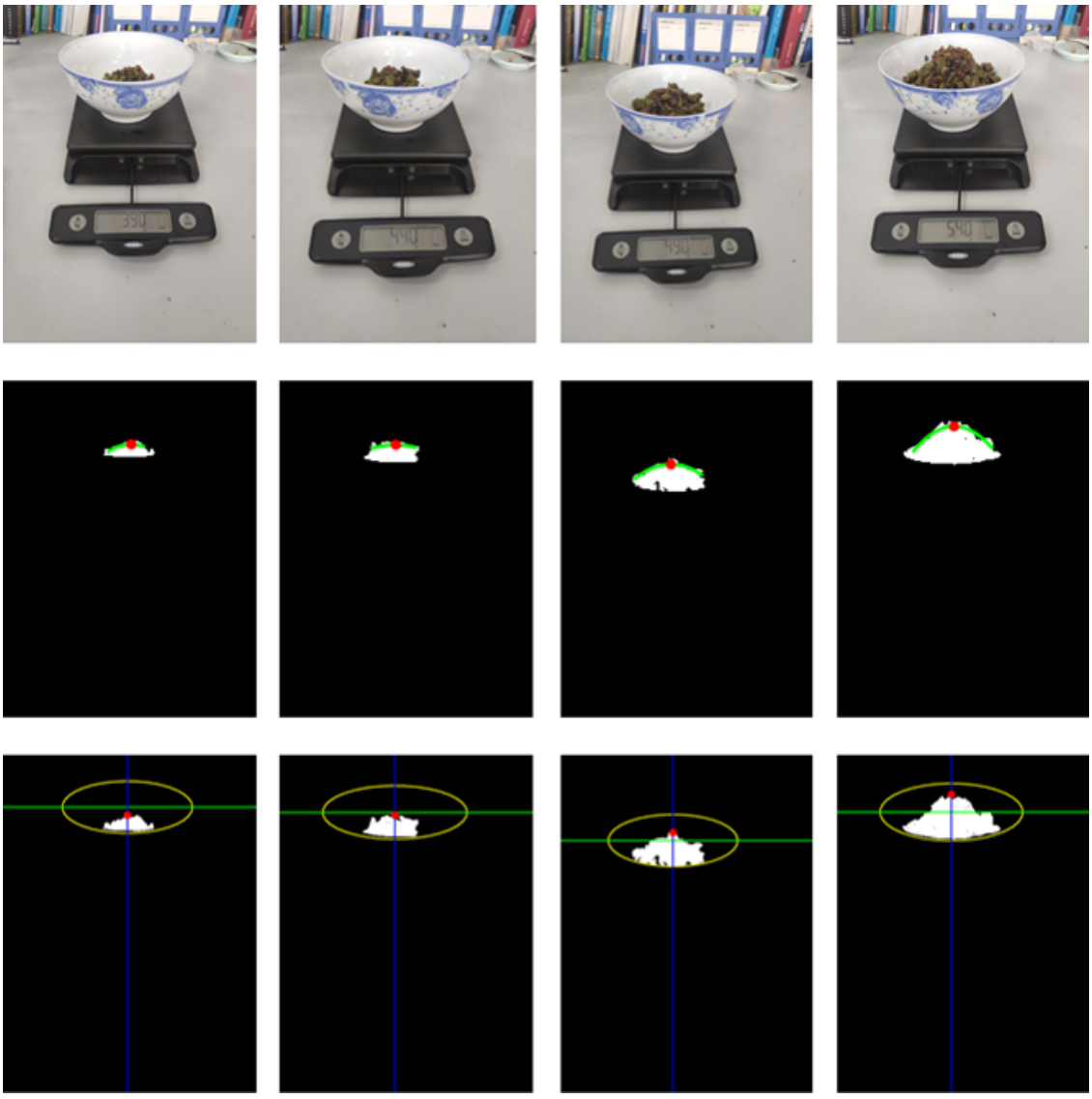
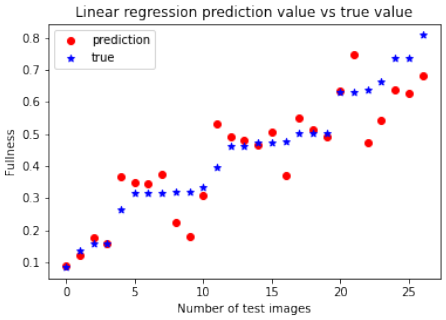
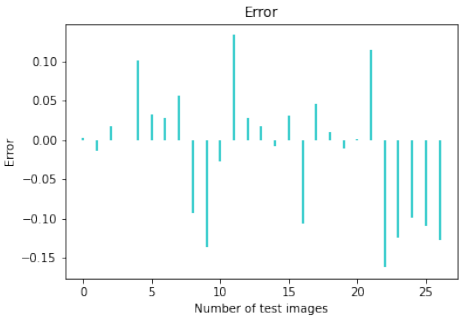


Figure 14: Normalized curvature/shape vertex effects on fullness.

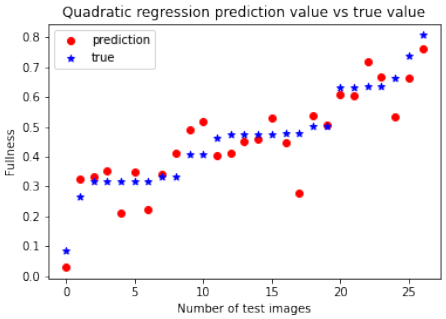
4.2 Regression



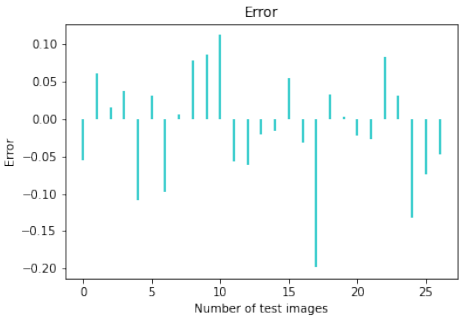
(a) Linear regression test result.



(b) Linear regression error.



(c) Quadratic regression test result.



(d) Quadratic regression error.

Figure 15: Regression model test result.

In total we have processed 135 different 2D images, and we perform 5-fold cross validation to evaluate the performance of the regression model. Fig 15 shows the test results for one of the five attempts. The results include the error and the comparison between the prediction value and the test value. Mean absolute error of linear regression model within the 5-fold cross validation is 0.061, which is similar to that of the quadratic regression model (0.059). Overall, linear regression has the similar performance as quadratic regression.

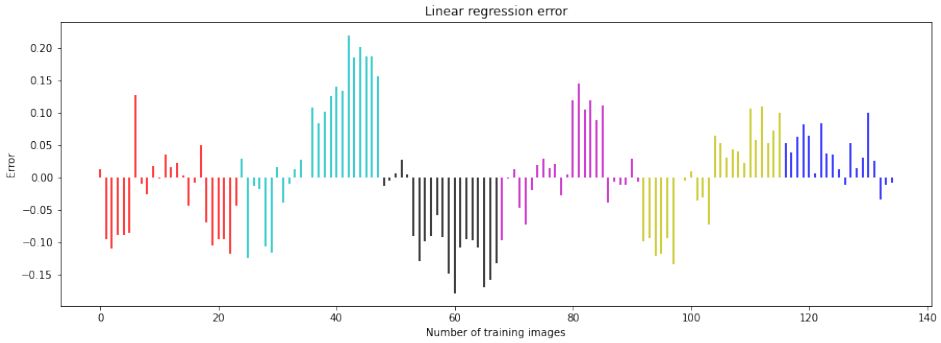
4.3 Discussion

Fig 12 shows the features extracted from a sample input image. According to the definition of normalized curvature and normalized shape vertex, we have noticed that sometimes these features won't be able to accurately demonstrate the property of the bulge of the food. Figure 12d shows the maximum curvature of the boundary between food and the side of the bowl. In this case, curvature does not help us to demonstrate the bulge of the food. When the bulge of the food in the bowl is not obvious or the detected bulge is not the actual bulge of the food, a large error may occur for this feature. Similarly, Figure 12e shows the shape vertex projection of the interaction between food and the side of the bowl, which also cannot achieve the goal to estimate the bulge of the food. But in this scenario, normalized shape vertex and normalized curvature can be used to demonstrate the food volume directly. Fig 14 shows the scenario when maximum curvature and shape vertex projection could help to demonstrate the bulge of the food. When the shooting angle is flat, the upper food contour clearly shows the bulge of the food. Otherwise, the upper food contour usually shows the interaction between the food and the side of the bowl.

Our model uses 4 features to estimate the food fullness and could obtain acceptable results. However, in order to better understand which features contribute more to the model, we have conducted more experiments. We reduced the number of features and then observe the changes in the regression model results. We found that when measured food orientation and food area ratio are removed, the results of the model will become very poor. We believe that these two features contribute the most to the regression model. According to Fig 13, we can see food area ratio share the same trend with the Fullness for a single kind of food. When we use food area ratio only in our feature set, we could produce a reasonable result but not as good as the result with 2 features (food orientation and food area ratio). According to our experiment, the ranking for the features would be: food area ratio, food orientation, normalized curvature/shape vertex and every feature contributes to the performance of the regression model.

Fig 16 shows the five-fold result for the proposed regression models. Each color represents a different food type. In total we have 135 2D images from 6 different kinds of food. As

we can see from the result, the error for a single kind of food usually goes into the same direction, either negative or positive. This means the regression model results are largely depending on the food types. Different food types may contribute to the performance of the regression machine learning model. If we want to improve the regression model, we may need to take food internal characteristics into consideration.



(a) Linear regression five-fold validation error.

(b) Quadratic regression five-fold validation error.

Figure 16: Regression model five-fold validation result.

5.0 Conclusions and Future Work

5.1 Conclusions

In this thesis, we have introduced a novel method to estimate the amorphous food (Chinese cuisine) volume in a circular bowl based on geometric features detected from a single image. Four features, food orientation, food area ratio, normalized curvature and normalized shape vertex were defined and extracted in this work. Linear and quadratic regression models were selected to estimate the food volume based on input feature set. The mean absolute error was less than 0.085 while using 135 input food pictures. This represents a significant improvement over the current self-reporting method which is inaccurate and unreliable.

This work focuses on the food volume estimation with a circular bowl instead of a plate as the container. Bowl is the most commonly used container in eastern cuisine and the shade of the side of the bowl on food may bring a lot of difficulties on the 3D reconstruction step. Also, amorphous food without clear pre-defined shape makes the 3D reconstruction step extremely hard. This work also focused on the bulge of the food which is the most important characteristic of the amorphous food. This work proposed the method to only extract the geometric features from a single 2D image without the need of 3D reconstruction and achieved good results.

5.2 Future Work

This work mainly focuses on the amorphous food volume estimation based on geometric features detected from a single image. According to the discussion in Chapter 4, we have found that the regression result is significantly affected by the type of the food, which means we may need some food internal features to distinguish the difference of each food, such as food texture. Also, normalized curvature and normalized shape vertex, which are the features designed to demonstrate the bulge of the food, contribute not much to our current

regression model. We may need to come up with more effective features to demonstrate the bulge of amorphous food since this is the most obvious property from the visual perspective.

Circular bowl is used as the food container throughout this work. Based on the predefined food container shape, we adopt adaptive Hough transform to extract the ellipse bowl contour within the 2D image. Most of our features were extracted based on the ellipse shaped bowl contour. In order to increase the flexibility of the research, arbitrary bowl shape could be introduced into the future work of the amorphous food volume estimation.

Bibliography

[1] Curvature and Radius of Curvature. Math24. (`https://www.math24.net/curvature-radius`)[Accessed on: February 15, 2021].

[2] Y. Song and H. Yan. Image segmentation algorithms overview. *CoRR*, abs/1707.02051, July 2017.

[3] X. Chang, H. Ye, P. Albert, D. Edward, K. Nitin, and B. Carol. Image-based food volume estimation. *CEA'13 : proceedings of the 5th International Workshop on Multimedia for Cooking & Eating Activities : October 21, 2013, Barcelona, Spain. Workshop on Multimedia for Cooking and Eating Activities (5th : 2013 : Barcelona, Spain)*, 2013:75–80, October 2013.

[4] F. Zhu, M. Bosch, I. Woo, S. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE Journal of Selected Topics in Signal Processing*, 4(4):756–766, August 2010.

[5] H. C. Chen, W. Jia, Z. Li, Y. N. Sun, and M. Sun. 3D/2D model-to-image registration for quantitative dietary assessment. In *2012 38th Annual Northeast Bioengineering Conference (NEBEC)*, pages 95–96, March 2012.

[6] Dietary Assessment Primer, 24-hour Dietary Recall (24HR) At a Glance. National Institutes of Health, National Cancer Institute. (`https://dietassessmentprimer.cancer.gov/`)[Accessed on: February 25, 2021].

[7] Dietary Assessment Primer, Food Record at a Glance. National Institutes of Health, National Cancer Institute. (`https://dietassessmentprimer.cancer.gov/`)[Accessed on: February 25, 2021].

[8] Dietary Assessment Primer, Food Frequency Questionnaire at a Glance. National Institutes of Health, National Cancer Institute. (`https://dietassessmentprimer.cancer.gov/`)[Accessed on: February 25, 2021].

[9] M. Haftenberger, T. Heuer, C. Heidemann, F. Kube, C. Krems, and G. B. Mensink. Relative validation of a food frequency questionnaire for national health and nutrition monitoring. *Nutrition Journal*, 9(1), September 2010.

[10] A. F. Subar. Developing dietary assessment tools. *Journal of the American Dietetic Association*, 104(5):769–770, May 2004.

[11] B. M. Silva, I. M. Lopes, J. J. P. C. Rodrigues, and P. Ray. Sapofitness: a mobile health application for dietary evaluation. In *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, pages 375–380, June 2011.

[12] M. A. Subhi, S. H. Ali, and M. A. Mohammed. Vision-based approaches for automatic food recognition and dietary assessment: a survey. *IEEE Access*, 7:35370–35381, March 2019.

[13] F. Zhu, A. Mariappan, C. J. Boushey, D. Kerr, K. D. Lutes, D. S. Ebert, and E. J. Delp. Technology-assisted dietary assessment. In Charles A. Bouman, Eric L. Miller, and Ilya Pollak, editors, *Computational Imaging VI*. SPIE, February 2008.

[14] I. Woo, K. Otsmo, S. Kim, D. S. Ebert, E. J. Delp, and C. J. Boushey. Automatic portion estimation and visual refinement in mobile dietary assessment. *Proc SPIE Int Soc Opt Eng*, 7533, January 2010.

[15] F. Kong and J. Tan. Dietcam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, 8(1):147–163, February 2012.

[16] S. M. Dimitratos, J. B. German, and S. E Schaefer. Wearable technology to quantify the nutritional intake of adults: validation study. *JMIR mHealth and uHealth*, 8(7):e16405, July 2020.

[17] M. L. Magrini, C. Minto, F. Lazzarini, M. Martinato, and D. Gregori. Wearable devices for caloric intake assessment: state of art and future developments. *The Open Nursing Journal*, 11(1):232–240, October 2017.

[18] M. Sun, J. D. Fernstrom, W. Jia, S. A. Hackworth, N. Yao, Y. Li, C. Li, M. H. Fernstrom, and R. J. Sclabassi. A wearable electronic system for objective dietary assessment. *J Am Diet Assoc*, 110(1):45–47, January 2010.

[19] eButton. Laboratory for Computational Neuroscience, Departments of Neurosurgery, Electrical and Computer Engineering, Bioengineering, University of Pittsburgh. (`http://lcn.pitt.edu/ebutton/`)[Accessed on: January 25, 2021].

[20] J. Shang, M. Duong, E. Pepin, Xing Zhang, K. Sandara-Rajan, A. Mamishev, and A. Kristal. A mobile structured light system for food volume estimation. In *2011*

IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pages 100–101, November 2011.

[21] W. Jia, Y. Yue, J. D. Fernstrom, Z. Zhang, Y. Yang, and M. Sun. 3D localization of circular feature in 2D image and application to food volume estimation. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4545–4548, August 2012.

[22] L. S. Davis, A. Rosenfeld, and J. S. Weszka. Region extraction by averaging and thresholding. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-5(3):383–388, May 1975.

[23] X. Yu, S. Xu, L. Jin, and E. Song. Characteristic analysis of Otsu threshold and its applications. *Pattern Recognition Letters*, 32(7):956–961, May 2011.

[24] T. Chaira. *Medical Image Processing: Advanced Fuzzy Set Theoretic Techniques*. CRC Press, Taylor & Francis Group, 2015.

[25] J. F. Haddon. Generalised threshold selection for edge detection. *Pattern Recognition*, 21(3):195–203, July 1988.

[26] D. Lin, J. Dai, J. Jia, K. He, and J. Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation, April 2016.

[27] G. Papandreou, L. Chen, K. Murphy, and A. L. Yuille. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation, Feburary 2015.

[28] G. Kumar and P. K. Bhatia. A detailed review of feature extraction in image processing systems. In *2014 Fourth International Conference on Advanced Computing Communication Technologies*, pages 5–12, Feburary 2014.

[29] P V.C. Hough. Method and means for recognizing complex patterns. December 1962.

[30] R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, January 1972.

[31] HK Yuen, J Princen, J Illingworth, and J Kittler. Comparative study of hough transform methods for circle finding. *Image and Vision Computing*, 8(1):71–77, June 1990.

[32] E. Ostertagová. Modelling using polynomial regression. *Procedia Engineering*, 48:500–506, November 2012. Modelling of Mechanical and Mechatronics Systems.