Smartphone-Based Food Weight and Calorie Estimation Method for Effective Food Journaling

Elder Akpro Hippocrate AKPA*, Hirohiko SUWA*, Yutaka ARAKAWA*, and Keiichi YASUMOTO*

Abstract: World health statistics about overweight and obesity show that overweight, obesity and diet-related diseases still remain major health risks. According to the World Health Organization, most of the world’s population live in countries where overweight and obesity kill more people than underweight. Recently, many studies have been driven by the motivation of elaborating the “ideal” solutions to prevent and/or monitor overweight, obesity and diet-related diseases, and to encourage healthy diet and lifestyle. In this work, we present our idea to automatically measure food weight and calories, from food photo using ordinary chopsticks as a measurement reference. The analysis of the obtained results show that the use of near-by eating utensils combined with computer vision techniques is a great and exploitable approach to ubiquitously help in diet assessment and obesity treatment.

Key Words: calorie estimation, food weight measurement, eating tools, image processing, food journal.

1. Introduction

Overweight and obesity are described as an abnormal and/or excessive fat accumulation that may cause prejudice to one’s health. The World Health Organisation (WHO) revealed in its recent update Fact Sheet (June 2016) that 39% of the world adults aged over 18 years old were overweight, and 13% were obese in 2014 [1]. Medical studies revealed that overweight and obesity are generally the result of energy imbalance between the amount of energy (calorie) coming from consumed foods and the amount of calories burned by the body for daily life activities (breath, walk, speak,...). The clinical treatments and prevention of overweight and obesity require the patients to measure, record, and control daily food intake. The daily recorded food intake is then compared against the body-burned energy, to provide sufficient feedback to the patients. This approach is known as self-report diet tracking or food journaling.

However, many studies pointed out that self-report is not reliable and should not be used as a measure of true energy intake [2],[3]. Self-reports are mostly inaccurate for diet assessment due to reasons such as the unwillingness to report, the lack of nutrition information, the inability to estimate the amount of food, and the annoying feelings of writing down everything that has been eaten [4]. Despite the limitations observed with self-reports dietary intake, we can not exclude this approach in the treatment of overweight and obesity, because they contain valuable, rich, and critical information about foods and beverages consumed by populations that can be used to inform nutrition policy and assess diet [2]. There is still a need to support patient self-reports, and researchers believe that this support can be achieved by employing technological innovation [5]. The technology progress of recent years in the area of digital cameras and the ubiquitous property of smartphones are perceived by the research community as an opportunity to improve diet assessment, as well as a mean to track and monitor food intake beyond clinical boundaries. Indeed, when it comes to eating activity tracking and monitoring, the main sensor utilized is the user’s phone camera. In many studies, food images taken with users’ cameras are seen as a paramount input to assess foods nutrients. For example, Kawano et al. [6] proposed a mobile food recognition system for estimating calorie and nutrition of foods, and records users’ eating habits. The system is a real-time recognition system that runs image recognition techniques on an Android device, using the device embedded camera. In another system [7], Ye et al. introduced a dietary assessment system that records daily food intake through the use of food photo taken at a meal time. In a similar work entitled Im2Calories [8], authors presented a system which can recognize the contents of a meal from a single image, and then predict its nutritional contents, such as calories. They assume that the user is eating at a restaurant for which they know the menu.

Despite the existence of many image-based food nutrient calculation systems and research studies, we have to admit that, they generally focus on food identification and classification. Few studies have been done on automated food weight measurement from a single image, because of the challenges of obtaining accurate weight from only one image [9]. Up-to-date image-based food weight estimation systems require the user to take multiple images or use markers, to graphically reconstruct the food in the images. Food weight estimation from a single image still remains an open and challenging problem to overcome for building highly acceptable calorie and nutrient estimation system; because without knowledge of the food weight, we can not estimate the amount of calorie, needed in the treatment of diet-related diseases [10].

In this paper, we propose a novel way to estimate food weight and calorie from a single image, with the help of chopsticks used during meal times, for eating. We use chopsticks of length 22 cm of our university’s restaurant, as measure reference, and
apply computer vision techniques on the food picture, which contains the chopsticks. The fact that the system uses daily-life chopsticks, which can be found everywhere, makes the system pervasive and suppress the burden of always carrying calibration objects as observed in existing systems like in [7].

The proposed system just requires the user to take a single image from the top with chopsticks visible in the image as shown in Fig. 1, with one stick positioned on the edge of the container, while the other stick is put on the table close to the container. Via a smartphone application, the user sends the taken image to the server. At the server side, the system automatically measures the diameter and the height of the food container using several images processing techniques, the camera’s focal length and sensor size contained in the exchangeable image file (EXIF) metadata of the image. Then, given the food type, the system combines the information about the container diameter, the height and the food type to provide the weight of the food in the image, and finally estimates the calories of the food. Our proposed method aims to estimate a calorie of a food by:

- calculating its weight from a single shot image,
- estimating the calorie from the calculated weight using a weight to calorie nutrient database tables.

In an experimental trial ran over 15 food types images collected in our university’s restaurant, the system achieved an average relative error rate of 6.65% for the weight measurement and 6.70% relative error for the calorie estimation. In addition, our proposed system constructs a personal food journal that records daily consumed meals, through the developed Android application. We present two contributions to the research field of image-based calorie estimation:

- Usage of eating tools as measurement reference, in replacement of the usually use calibration cards,
- A calorie estimation method that considers foods served in containers having bowl shape.

2. Related Work

In this section, we discuss existing systems and methods for calorie assessment in terms of image-based systems and volume-weight estimation systems that use computer vision techniques.

2.1 Related Food Image-Based Calorie Assessment Systems

There has been some research on estimating dietary composition of meals using image analysis. Woo et al. [11] and Zhu et al. [9] proposed a Technology-Assisted Dietary Assessment (TADA) system to process food images with a mobile device. However, in that system it is assumed that the plate is white, food items in the plate are separated, and users have to take food photos with a chessboard-like marker to calibrate the images; all of these settings are hardly possible in real eating environment. The interesting point of this work is setting users free from any extra operations besides shooting pictures through automated food recognition and 3D volume reconstruction. As a food recording system with food recognition function, FoodLog [12], [13] estimates food balance by dividing a received food image into 300 blocks. For each blocks, it extracts the color and discrete cosine transform (DCT) coefficients, and classifies into five groups such as staple, main dish, side dish, fruit, and non-food. Yang et al. [14] proposed a method to estimate the nutrition in foods. They calculate pairwise statistics between local features computed over pixel-level segmentation of the food image into eight ingredient types. The accuracy of this system was up to 28%.

In the above-mentioned systems, the work is mainly focused on the identification of the “menu” or the “food item”, without much consideration of the food volume or the food weight in their estimation. Our proposed system can estimate the food weight with no marker required. Note that our current system estimates only food weight and not food types, and it requires user’s assistance to enter the food name.

2.2 Related Food Quantity Estimation Systems

Food portion size estimation is extremely difficult since foods shapes and appearances are subject to variations due to the cooking circumstances and eating conditions. For this reason, food volume and weight estimation for dietary management have been a challenge for ubiquitous computing and healthcare systems.

Most image-based food volume estimation systems use multiple images [15],[16], video [17] or 3D reconstruction [18]. For example, Kong et al. developed an application called iEatCam [19] that automatically assesses food intake based on multiple views of food. The user is required to take three pictures of the same food separated by 120 degrees, in order to get the food volume. This requirement increases the burden on the user. Another approach appears in [20] where two pictures must be taken, one from the top and other from the side, with the user’s thumb placed beside the dish when taking the picture from the top. In this study, they explored three categories of food: single food, non-mixed food, and mixed food. While this study showed good results for single and non-mixed foods such as eggs, oranges, and apples, it had problems with mixed foods such as soup and curry. Chen et al. [21] introduced a 3D/2D model-based image registration method for quantitative food intake assessment. Their method uses global contours to determine the position, the orientation, and the scale of the user-selected 3D shape model. The volume estimation obtained is accurate for food items such as oranges or hamburgers with a simple model. However for food items such as bananas or salads that have a complex model, this system does not have a
Compared with the above methods, our proposed system is less obtrusive, less burdensome, suitable for any eating-environment and requires only one food picture for food weight and calorie measurement.

3. Proposed Food Weight Measurement Method

The goal of our work is to help users record daily food and obtain an accurate weight estimation from a single picture of the food taken before eating. An important part of this process is the use of chopsticks in the food image. Chopsticks, which can be found easily in any kind of eating environments such as homes, restaurants, and vending machines, are used as a reference for measurement. We take advantage of their industry standards and well-known length 20 cm to 22 cm to compute the volume and weight of the foods in the images. In this work, we utilized disposable chopsticks measuring 22 cm, which are available in the university’s restaurant. We also used the fact that generally foods are served in standardized food containers, designed with regular shapes and normalized dimensions. Thanks to this standard, the Japanese curry rice, for example, is rarely served in containers for ramen (Chinese noodle).

In the early development stage of our system, we used chopsticks as reference measurement to only estimate the diameter of the container in the picture and then derive the height of the container from a pre-built diameter-to-height table, which matches the diameter to the height based on the food type. However, this approach was highly dependent on the accuracy of the pre-built diameter-to-height table, and the obtained result was not satisfactory enough. Thereupon, we proposed a weight estimation method that computes the diameter and the height of the food container from a single image of the food, using several computer vision techniques combined with the EXIF metadata of the image [22]. With the food name provided, the system is able to estimate the food volume. Then, we used the food volume and a density table to extract the mass of the food, and finally estimate the food caloric content. In this section, firstly, we will explain how we compute the diameter of the food container through image processing. Secondly, we will demonstrate how to get the container height from a single image, using chopsticks and finally, we will explain how we combine the food type, density table values and container sizes, to obtain the food volume and its weight.

3.1 Measurement of Food Container Diameter

To measure the diameter of the food container from the food picture, we need to know the geometric shape and the size (pixels) of the container and the length of the chopsticks (in pixels) in the picture.

3.1.1 Container shape detection

For the geometric shape problem, since foods are generally served in round shape containers, we decided to initially detect the circle shape of the containers open-top. To detect the circle shape of round dishes, we use a computer vision feature extraction technique called the Hough transform.

The Hough transform can be used to isolate features of a particular shape within an image. Because it requires the desired features to be specified in some parametric form, the classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc. [23]. The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise.

The Hough Transform has two essential parameters, m and n. The parameter “m” is used for the Canny edge detector, while “n” is for the center detection stage [24]. The Canny operator is an optimal edge detector that takes as input a gray scale image, and produces as output an image showing the positions of tracked intensity discontinuities, which are the edges of the input image. The number of circles detected by the Hough transform depends on these two parameters. To detect the circle of the main round dish with high precision, after several tests, we fixed these two parameters to 10 and 100. The smaller n is, the more false circles may be detected. However, with n fixed to 100, we always obtained good circle detection with all the tested food images. The result of the Hough transform is the circle formed by the round shape of the container and its radius in pixels.

We store this value of the radius, which will be used later with the result of the next subsection, to find the diameter of the container in real life.

3.1.2 Determination of the chopsticks length in the image

The next step consists of getting the length of the chopsticks in the image. We achieve this step by applying the Canny edge algorithm to find the edges of the input image (Fig. 2). The Canny edge requires three parameters: low threshold, high threshold, and the kernel size.

The low threshold is used for edge linking, while the high threshold is used to find a segment with strong edges. In the previous Section 3.1.1, we presented the parameter “m”. This parameter represents the low threshold that the Canny edge uses to link the edges. If no value is given to the high threshold, the value of the high threshold is, by default, set to three times the value of the low threshold, according to Canny’s recommendation [25].

The kernel size is the size of the Sobel operator to be used inside. The Sobel operator is an algorithm which emphasizes regions of high spatial frequency that correspond to edges.

The output obtained from the Canny algorithm is processed to find contours in the food image. From the possible contours found, we only extract rectangular contours and then retrieve the length in pixels of the chopsticks by taking the length of the stick placed on the container, which appears as the longest stick in the food image (Fig. 2).

Finally, all the obtained results, the container’ radius in the picture, the chopsticks length in the image and the known length of the chopsticks in real life, are combined in a cross multiplication to determine the diameter of the container in real life.

Table 1 shows the results of our diameter estimation method applied to 25 images of 5 different containers' types (5 images for each type of container). The pictures were taken at different positions but from the top, and the 25 containers were empty. We decide to run the first test on empty containers to evaluate the accuracy of the proposed method before testing the system with non-empty containers. The average estimation, the standard deviation (SD), and the relative error of each estimation are also presented in the table.

Figure 3 shows the five different types of bowl used for testing. “Small bowl” refers to bowl type used to serve white rice
Fig. 2 Processing steps for container diameter and chopsticks length measurement.

![Fig. 2 Processing steps for container diameter and chopsticks length measurement.](image)

(a) original image.  (b) edges detection.  (c) thresholded image.  (d) output image.

Fig. 3 Types of bowl used for testing.

![Fig. 3 Types of bowl used for testing.](image)

Table 1 Container diameter estimation.

<table>
<thead>
<tr>
<th></th>
<th>Miso bowl</th>
<th>Small bowl</th>
<th>Big bowl</th>
<th>coffee or tea cup</th>
<th>water or juice glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real diameter (cm)</td>
<td>12</td>
<td>12</td>
<td>16</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Picture 1</td>
<td>11</td>
<td>12</td>
<td>16</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Picture 2</td>
<td>11</td>
<td>11</td>
<td>16</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Picture 3</td>
<td>12</td>
<td>11</td>
<td>15</td>
<td>6.5</td>
<td>7</td>
</tr>
<tr>
<td>Picture 4</td>
<td>12</td>
<td>12</td>
<td>17</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Picture 5</td>
<td>10</td>
<td>11</td>
<td>15</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Average Estimation</td>
<td>11.20</td>
<td>11.40</td>
<td>15.40</td>
<td>7.30</td>
<td>6.40</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.83</td>
<td>0.57</td>
<td>1.14</td>
<td>0.70</td>
<td>0.54</td>
</tr>
<tr>
<td>Relative Error (%)</td>
<td>6.67</td>
<td>5.00</td>
<td>3.75</td>
<td>8.75</td>
<td>8.57</td>
</tr>
</tbody>
</table>
of the camera. The point $R$, referred as the image center, corresponds to the intersection of the image plane and the optical axis. The projection of the 3D point onto the 2D image plane is the point $Q$. This point lies at the intersection of the image plane and the projection line. The point $Q$ has coordinates $(x_i, y_i)$ in the 2D coordinate system $(X_i, Y_i)$ with the origin $R$ of the image plane.

To extract the distance from the camera to the point $P$, we need to find the mathematical relation between the 2D coordinates of the point $Q$ $(x_i, y_i)$ and the 3D coordinates $(x_w, y_w, z_w)$ of the point $P$.

### 3.2.2 Triangle similarity

Two geometrical objects are similar if each object is congruent to the result of a particular uniform scaling of the other object. Figure 6, which represents the same scene as Fig. 5, but from above, looking down in the negative direction of $Y_w$ axis, possess two similar triangles $O_c fQ$ and $O_c Z_w P$.

![Pinhole camera model seen from $Y_w$.](image)

Both triangles have part of the projection line as hypotenuses. The two adjacent sides of the left triangle to the right angle are $-x_i$ and $f$ (focal length of the camera). For the right triangle, the adjacent sides are $x_w$ and $z_w$. Since these two triangles are similar, we get the following equations:

$$
\frac{-x_i}{f} = \frac{x_w}{z_w} \Rightarrow x_i = -\frac{f}{z_w}x_w.
$$

(2)

In the same way, when looking in the negative direction of $X_w$ axis of the world coordinate, we get:

$$
\frac{-y_i}{f} = \frac{y_w}{z_w} \Rightarrow y_i = -\frac{f}{z_w}y_w.
$$

(3)

From these two equations, the relation between $(x_i, y_i)$ and $(x_w, y_w, z_w)$ is derived:

$$
\begin{bmatrix}
  x_i \\
  y_i
\end{bmatrix} = \begin{bmatrix}
  -\frac{f}{z_w} \cdot x_w \\
  -\frac{f}{z_w} \cdot y_w
\end{bmatrix}.
$$

(4)

Because the mapping of a point from the world coordinate to a 2D coordinate described in the pinhole model is a perspective projection followed by a 180 degree (180°) in the image plane, the resulting image is rotated 180 degree. In order to get the expected image from the camera, we need to rotate the coordinate system in the image plan by 180 degree. Finally, after rotation, the mapping from 3D coordinate of a point $P (x_w, y_w, z_w)$ to the 2D coordinate point $Q (x_i, y_i)$ is given by:

$$
\begin{align*}
  x_i &= \frac{f}{z_w} \cdot x_w \\
  y_i &= \frac{f}{z_w} \cdot y_w.
\end{align*}
$$

(5)

### 3.2.3 Measurement of the distance from camera to stick

To find the distance from a camera to an object or a marker, computer vision techniques exploit the triangle similarity and pinhole model results, described in the previous Sections 3.2.1 and 3.2.2. The camera to object distance is given by:

$$
\frac{P}{P_{\text{Distcam}}} = \frac{W}{\text{Dist}_{\text{cam, obj}}} = \frac{F \cdot W}{O_w \cdot S_z},
$$

(6)

where $F$ (mm) is the focal length of the camera, $W$ (mm) the known width of the object or marker, and $P$ (pixels) the apparent width of the object in the image.

Given the sensor size of the camera, the formula can be defined as:

$$
\text{Dist}_{\text{cam, obj}} = \frac{F \cdot W \cdot I_w}{O_w \cdot S_z},
$$

(7)

where $I_w$ is the image width, $O_w$ the object width (pixels) and $S_z$ the sensor size (mm).

In our current implementation, we use the food images from a smartphone and for each image we extract the EXIF data of the camera system, which contains metadata such as the camera focal length and the sensor size. In Section 3.1, we explained how to get the width of each stick of the chopsticks. We input these values into (7) to determine the distance $d_{CS1}$ and $d_{CS2}$ from the camera to each stick. Finally, we get the height $H$ of the food container by taking the difference between $d_{CS1}$ and $d_{CS2}$ (see (1)).

Table 2 shows the results of applying our height measurement method to the same 25 images of empty containers as described in Section 3.1.2. The table also gives the average estimation, the standard deviation (SD) along with the relative error of the measurement for each container.

As shown in the table, the height estimation method achieves a relative error rate of 1.33% in the best case with the “Miso bowl,” and 11.54% in the worst case, with the “Small bowl.”

### 3.3 Volume Estimation and Food Type Identification

Estimation of the food volume is a critical step for any calorie estimation system for dietary assessment. In our proposed system, after determining the container dimensions, we estimate the food volume from the food image by firstly estimating the container volume, and then using the density information for that particular food, we estimate its volume and weight. Currently, we use the “aqua-calc” table [26] which provides a volume to mass conversion for more than 4000 food items and ingredients.
3.3.1 Volume estimation

In our current implementation, we focus on the foods that are served in containers of bowl shape. A bowl is a round, open-top container used in many cultures, especially in Japanese culture, to serve foods. Geometrically, classic bowl can be represented as spherical cap. A spherical cap is a section of a sphere slide off by a plane (Fig. 7).

![Fig. 7 Spherical cap.](image)

The volume of a spherical cap is defined by this ordinary used equation:

\[ V = \frac{\pi}{6} h (3r^2 + h^2) \]  \hspace{2cm} (8)

where \( r \) is the radius of the base of the cap, and \( h \) its height. These two parameters are obtained from the methods described in the Sections 3.1 and 3.2.

Figure 8 shows the results of our volume estimation using (8). The best volume estimation are obtained with the “Water or juice glass”, and the“coffee or tea cup” with respectively a percentage error rate of 1.38% and 1.98%. The system produces acceptable results for other container types.

Knowing the container volume, which we assume to be closely equal to the food volume when the bowl is full, we can get the food weight using density table information of each food type.

3.3.2 Food type identification

Food recognition and classification has been the focus of many image-processing studies in the recent years. Among these studies, some exploited the local features of food images to identify the food type [14], while other focused on global features such as global histogram [27]. Other studies utilized features extraction techniques such as scale invariant feature transform (SIFT) [28] for food image classification. Kagaya et al. proposed a food detection and recognition system using a deep-learning approach called convolutional neural network (CNN) [29]. However in most cases, the accuracy of the classifiers proposed in these study are questionable and/or require further improvements to perform as expected.

In our current implementation, we do not implement any food classification technique; we manually identify the food type by entering the food name. Moreover, in our end-user system application, we intend the user to confirm the classified food, to reduce misclassification and improve the overall system accuracy.

As mentioned before, the most critical and challenging step in dietary assessment is the food weight estimation from the food image. Therefore, in this paper, we assume that the food type has been well identified or given manually.

3.3.3 Density table

Food density tables are databases which provide a tool for researchers and professionals of food analysis to convert volume into weight and vice-versa. Data collected are prepared from the literature, various national food composition tables and measurements are conducted by international and national organization such as FAO (Food and Agriculture Organization of the United Nations), FNDDS (Food and Nutrient Database for Dietary Studies) or USDA (United States Department of Agriculture).

In this study, we used the “aqua-calc” density [26], since it has data available for Japanese food types as well as international food types. It provides a volume to mass conversion for more than 4000 food items and ingredients. Table 3 shows for some foods, their density along with the conversion to weight and vice-versa. Data collected are prepared from the literature, various national food composition tables and measurements are conducted by international and national organization such as FAO (Food and Agriculture Organization of the United Nations), FNDDS (Food and Nutrient Database for Dietary Studies) or USDA (United States Department of Agriculture).

In our system, the food weight is obtained by multiplying the estimated volume acquired from the food image with the specific food density as shown in (9).

\[ \text{food weight} = \text{estimated volume} \times \text{food density} \]  \hspace{2cm} (9)

### 4. Food Journaling and Experiment

To verify the measurement obtained from our proposed estimation system, we performed validation experiments using images of various food types (rice, ramen, miso, oyakodon, gyudon) served in the five container types mentioned before. Over three weeks, we collected the food images of 15 participants (12 male) aged 23-31, through an Android food journal application (app) that we developed. To certify the accuracy of our system, from the collected images, we selected only images that are from foods served in our university restaurant, since we know their pre-estimated weight and calorie. The details about this food journal app and the results of our experiment are discussed in the following sections.

#### 4.1 Image Collection through Smartphone Application

One mission of the proposed system is to help people who suffers from obesity, overweight and diet-related disease to...
keep a record of the amount of daily nutrients they consumed without the need for recording this data manually. The functions of our system are to calculate not only the weight of the food in the picture, but also the amount of calories and to keep a food log of the users’ consumed foods. To accomplish this, we developed a smartphone app that runs on Android devices and takes advantage of the built-in camera, processing power and network capabilities of the smartphones. The app represents the interface for users to send and receive data about their foods. Over three weeks, the participants used this app to collect food images for testing the estimation method proposed in this study. The application is also utilized in another research study, in which we assess the consistency of meal-tracking and behavior change using a game-based approach [30]. Figure 9 shows some screenshots of the application relevant to the food images collection and the proposed estimation method.

![Fig. 9 Smartphone application for food images collection.](image)

4.2 Results of the Weight Measurement

During the period of food images collection, we received 119 images from the participants. Among these images, about 40 images were not taken from top or with the chopsticks not positioned as shown in Fig. 1. Some other images were of not foods served in round dish containers. These kinds of images were initially accepted because the image collection application is used for a separate study, which does not have restrictions on the food photo, but we do not include them in this evaluation. Figure 11 shows some of the images received during the 3-week food image collection campaign.

![Fig. 10 Examples of food images collected.](image)

![Fig. 11 Examples of food images not used.](image)

In the “Welcome screen,” users have the control buttons to navigate to the game screens, upload screens and to other options screens such as “food log-calendar” and pop-ups. The “Upload screen” allows users to take a photo or choose a photo from the gallery and contains a form where users can fill in additional information such as hunger level and mood. Figure 10 shows some of the images received during the 3-week food image collection campaign.

4.3 Calorie Estimation

Once we know the estimated food weight, we can move on to the final step of estimating the amount of calorie (energy) of the food in the image. To achieve this step, we rely on nutritional fact database (NFD) and food guide as reference to provide nutritional information. Previous studies show that NFD and food guide are important components to realize useful and successful calorie estimation system [17]. Details about nutritional values of various food types are stored in these databases and are available from national and international health organizations. In Japan, the Ministry of Health, Labour and Welfare (MHLW) and the Ministry of Agriculture, Forestry and Fisheries (MAFF) introduced the “Spinning food guide” as the food and nutrition reference tool to help people practice healthy eating. Figure 12 shows the Japanese spinning food guide accompanied by a chart that indicates the recommended daily servings for each food group.

![Fig. 12 Japanese spinning food guide.](image)

Our system currently uses such databases as reference to estimate the amount of calorie of the food in the received images. For this study, since we used images of food served in our institute’s restaurant, we adopted the restaurant pre-estimated food calorie database as ground truth data to evaluate our proposed system. Table 5 shows for each food type the real values
Table 4 Results of proposed method in comparison with real values.

<table>
<thead>
<tr>
<th>Food type</th>
<th>Average measured weight (g)</th>
<th>Estimated weight (g)</th>
<th>Absolute Error</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oyakodon</td>
<td>468.5</td>
<td>450.5</td>
<td>-18</td>
<td>3.84</td>
</tr>
<tr>
<td>Tamagodon</td>
<td>444.5</td>
<td>422</td>
<td>-22.5</td>
<td>5.06</td>
</tr>
<tr>
<td>Katsudon</td>
<td>413.6</td>
<td>398.7</td>
<td>-14.9</td>
<td>3.60</td>
</tr>
<tr>
<td>Kakeudon</td>
<td>405</td>
<td>398.5</td>
<td>-6.5</td>
<td>1.60</td>
</tr>
<tr>
<td>Rice (small size)</td>
<td>160</td>
<td>174.5</td>
<td>14.5</td>
<td>9.06</td>
</tr>
<tr>
<td>Rice (regular size)</td>
<td>340.4</td>
<td>305.4</td>
<td>-35</td>
<td>10.28</td>
</tr>
<tr>
<td>Ramen</td>
<td>650</td>
<td>680</td>
<td>30</td>
<td>4.62</td>
</tr>
<tr>
<td>Miso soup</td>
<td>143</td>
<td>155</td>
<td>12</td>
<td>8.59</td>
</tr>
<tr>
<td>Gyudon</td>
<td>416.5</td>
<td>430.3</td>
<td>33.8</td>
<td>8.12</td>
</tr>
<tr>
<td>Fried rice</td>
<td>371.8</td>
<td>400</td>
<td>28.2</td>
<td>7.58</td>
</tr>
<tr>
<td>Tendon</td>
<td>380</td>
<td>400</td>
<td>20</td>
<td>5.26</td>
</tr>
<tr>
<td>Kake soba</td>
<td>375</td>
<td>402</td>
<td>27</td>
<td>7.20</td>
</tr>
<tr>
<td>Kitsunu soba</td>
<td>428</td>
<td>385.5</td>
<td>-42.5</td>
<td>9.93</td>
</tr>
<tr>
<td>Tsukimi soba</td>
<td>434</td>
<td>390</td>
<td>-44</td>
<td>10.14</td>
</tr>
<tr>
<td>Tonjiru</td>
<td>176</td>
<td>185</td>
<td>9</td>
<td>5.11</td>
</tr>
<tr>
<td><strong>Average relative error</strong></td>
<td><strong>6.65%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The average relative error for the calorie estimation is about 6.70%. The worst case of underestimation is observed from the Rice (regular size) with an absolute error of −59 kcal, while the worst overestimated food shown an absolute error of +63 kcal (gyudon). According to the USDA nutrient Database [32], 63 kcal represent the quantity of energy (calorie) in 1 medium raw egg of 44 g. From this, we can assert that our proposed system has an overall acceptable result, and can work better if we overcome some current limitations, described in the next section.

5. Discussion

Our experiment revealed several strengths and weaknesses in our proposed calorie estimation system. In the rest of this section, we discuss some of the strengths, limitations and how we plan to overcome them in our future work.

5.1 Food Container Type

To the best of our knowledge, this is the first study of food weight and calorie estimation that considers food served in bowls. In our proposed system, we focused on food served in bowl shape containers, but the methods explained can be applied to plate containers as well as other types of containers. Other existing works [9] required the use of plate container with a specific color. For example, [11] assumed that the food is served in a white plate and the food items in the plate should be separated. With respect to their works and results, in this study, we tried to explore a different aspect with foods served in bowls, from a single image, without the use of calibration items.

From the experiments, we found that our weight estimation has some limitations coming from the camera angle (angle of view and angle of shoot). When the food image is not properly taken from the top, the system fails to detect the container shape and the chopsticks in the image (Fig. 13).

To solve this problem, we will try to get the camera angle by either extracting more data from the EXIF metadata, or get the angle from the smartphone accelerometer sensor.

5.2 Food Serving Style and Containers

In this work, we assumed that foods are served to fill the containers. However, in real-life, foods are rarely served to fill the entire volume of the containers. There is almost always a space between the open-top of the container and the foods. This fact is probably the cause of over-estimation observed during the experiments. Another issue of the proposed system appears in the situation where a food menu is composed of many dishes or having one container for each food item of the menu. In such a scenario, the system is limited and cannot estimate the weight and calorie of all the dishes by using a single image. In this case, our target is only the main dish served in the largest bowl container. To estimate the calorie of the entire “set-menu”, the system needs to get the image of each dish of the menu and process them separately. However, this approach might be tedious for a menu with more than two dishes.
Table 5 Result of the food calorie estimation.

<table>
<thead>
<tr>
<th>N</th>
<th>Food Type</th>
<th>Weight (g)</th>
<th>Calorie (kcal)</th>
<th>Weight (g)</th>
<th>Calorie (kcal)</th>
<th>Absolute Error</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oyakodon</td>
<td>468.5</td>
<td>684</td>
<td>450.5</td>
<td>658</td>
<td>-26</td>
<td>3.80</td>
</tr>
<tr>
<td>2</td>
<td>Tamagodon</td>
<td>444.5</td>
<td>636</td>
<td>422</td>
<td>604</td>
<td>-32</td>
<td>5.03</td>
</tr>
<tr>
<td>3</td>
<td>Katsudon</td>
<td>413.6</td>
<td>823</td>
<td>398.7</td>
<td>794</td>
<td>-29</td>
<td>3.52</td>
</tr>
<tr>
<td>4</td>
<td>Kakeudon</td>
<td>405</td>
<td>320</td>
<td>398.5</td>
<td>315</td>
<td>-5</td>
<td>1.56</td>
</tr>
<tr>
<td>5</td>
<td>Rice (small size)</td>
<td>160</td>
<td>269</td>
<td>174.5</td>
<td>294</td>
<td>+25</td>
<td>9.29</td>
</tr>
<tr>
<td>6</td>
<td>Rice (regular size)</td>
<td>340.4</td>
<td>572</td>
<td>305.4</td>
<td>513</td>
<td>-59</td>
<td>10.31</td>
</tr>
<tr>
<td>7</td>
<td>Ramen</td>
<td>650</td>
<td>553</td>
<td>680</td>
<td>578</td>
<td>+25</td>
<td>4.52</td>
</tr>
<tr>
<td>8</td>
<td>Miso soup</td>
<td>143</td>
<td>46</td>
<td>155</td>
<td>50</td>
<td>+4</td>
<td>8.70</td>
</tr>
<tr>
<td>9</td>
<td>Gyudon</td>
<td>416.5</td>
<td>771</td>
<td>450.3</td>
<td>834</td>
<td>+63</td>
<td>8.17</td>
</tr>
<tr>
<td>10</td>
<td>Fried rice</td>
<td>371.8</td>
<td>673</td>
<td>400</td>
<td>724</td>
<td>+51</td>
<td>7.58</td>
</tr>
<tr>
<td>11</td>
<td>Tendon</td>
<td>380</td>
<td>642</td>
<td>400</td>
<td>676</td>
<td>+34</td>
<td>5.30</td>
</tr>
<tr>
<td>12</td>
<td>Kakesoba</td>
<td>375</td>
<td>304</td>
<td>402</td>
<td>326</td>
<td>+22</td>
<td>7.24</td>
</tr>
<tr>
<td>13</td>
<td>Kitsun no soba</td>
<td>428</td>
<td>441</td>
<td>385.5</td>
<td>398</td>
<td>-43</td>
<td>9.75</td>
</tr>
<tr>
<td>14</td>
<td>Tsukimi soba</td>
<td>434</td>
<td>378</td>
<td>390</td>
<td>340</td>
<td>-38</td>
<td>10.05</td>
</tr>
<tr>
<td>15</td>
<td>Tonjiru</td>
<td>176</td>
<td>121</td>
<td>185</td>
<td>128</td>
<td>+7</td>
<td>5.79</td>
</tr>
</tbody>
</table>

Average relative error 6.70%

5.3 Edge Detection Limitation

The shape detection steps can sometimes fail to detect the edge of the containers and the chopsticks. These detection failures can be caused by various environmental parameters such as the lighting conditions, illuminations and the color of the eating tables or containers [20]. In our work, all the photos were taken by the participants in their usual and natural eating environments, without any environmental bias. The use of such images taken in these non-pre-defined eating environments, with no control on the illuminations, can affect the result of the shape detection mechanism. However, despite the use of these food photos taken without any environmental bias, the proposed system achieved a good edge detection rate of approximately 64%. Indeed, out of the 79 properly taken photos (described in Section 4.2), the system succeeded to correctly detect the edges in 50 images used for the evaluation of the overall system. We believe that this rate of edge detection without environment bias is acceptable to achieve an effective food journaling system.

6. Conclusion

In this study, we proposed a food weight measurement and calorie estimation system based on image processing, which uses a smartphone camera and chopsticks as measurement reference. The system, also, constructs a food journal to keep track of daily consumed meals. We exploit image-processing techniques and the EXIF metadata of the food image (camera focal length, sensor size) to measure the food container size, determine the food volume, get the food weight and estimate the calorie by using density and nutrient database information of that particular food. An important aspect of this process is the use of chopsticks, which suppress the obligation of carrying and using calibration objects utilized in others systems. The conducted experiments show tenable results from the system which achieved an average relative error rate of 6.65% for the weight measurement, and 6.70% relative error rate for the calorie estimation. We believe that the proposed method could be used as helping tool for use in treatment of obesity, overweight and diet-related disease.

To provide more accurate and useful tool for dietary assessment, we intend to improve this system by taking into account the distortion coefficient of each smartphone camera lens, and extending our work to the use of other cutlery such as spoon and fork. Moreover, to enhance the food container detection, we plan to get the camera view angle from the EXIF metadata or the smartphone’s accelerometer sensor, when the photo is taken.

Acknowledgments

This work was partly supported by JSPS Kakenhi Grant Number 16H01721.

References


Elder A. H. AKPA
He received his M.E. degrees from Nara Institute of Science and Technology (NAIST), Japan, in March 2016. He is currently a Ph.D. student at NAIST University. His research interests include IT healthcare, wellbeing, behaviour change and diet tracking systems.

Hirohiko SUWA
He has been an assistant professor of Ubiquitous Computing Lab, in Nara Institute of Science and Technology (NAIST) since 2014. His current research interests are social informatics, real data mining and social network analysis. He is a member of IEICE, IPSJ, IEEE.

Yutaka ARAKAWA
He received the B.E., M.E., and Ph.D., from Keio University, Japan, in 2001, 2003, and 2006, respectively. He was an assistant professor at Keio University in 2006, and became an assistant professor at Kyushu University in 2009. His current research interests are human activity recognition, behavior change, and location-based information system. He is a member of IEEE, ACM, IPSJ, and IEICE.

Keiichi YASUMOTO (Member)
He received the B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University, Osaka, Japan, in 1991, 1993 and 1996, respectively. He is currently a professor of the Graduate School of Information Science at Nara Institute of Science and Technology. His research interests include distributed systems, mobile computing, and ubiquitous computing. He is a member of ACM, IEEE, IPSJ and IEICE.