SCalE: Smartphone-based Calorie Estimation From Food Image Information

Nafis Sadeq, Fazle Rabbi Rahat, Ashikur Rahman
Department of Computer Science and Engineering,
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Email: {nafissadeq, fr.rahat}@gmail.com, ashikur@cse.buet.ac.bd

Sheikh I. Ahamed, Md Kamrul Hasan
Department of Math., Stat. and Comp. Sci.
Marquette University
Milwaukee, USA
Email: {sheikh.ahamed,mdkamrul.hasan}@mu.edu

***Abstract*—Personal assistive systems for diet control can play a vital role to combat obesity. As smartphones have become inseparable companions for a large number of people around the world, designing smartphone-based systems is perhaps the best choice at the moment. Using this system people can take an image of their food right before eating, know the calorie content based on the food items on the plate. In this paper, we propose a simple method that ensures both user flexibility and high accuracy at the same time. The proposed system employs capturing food images with a fixed posture and estimating the volume of the food using simple geometry. The real world experiments on different food items chosen arbitrarily show that the proposed system can work well for both regular and liquid food items.**

***Index Terms*—Reducing obesity, food image, calorie estimation**

I. Introduction

World Health Organization (WHO) marks obesity as a significant risk for cardiovascular diseases, diabetes, musculoskeletal disorders, some cancers (including endometrial, breast, ovarian, prostate, liver and colon), etc. It has a severe impact on children. Obese children experience breathing difficulties, increased risk of fractures, hypertension, early markers of cardiovascular disease and insulin resistance. WHO also marks the fundamental cause of obesity and overweight which is 'an energy imbalance between calories consumed and calories expended' [1]. Thus, one of the ways of reducing obesity is to keep calorie intake under control.

As concern for health is growing day by day, different personal assistive tools for keeping good health are also receiving considerable attention. Nowadays, smartphones have become inseparable companions for the people all around the world. With the improvement of the capabilities of smartphones, different high processing tasks are possible with them. As a result, it has become easy to put different automatic tools on smartphones to aid the user in various fields [2]–[8]. Automatic food recognition and estimating nutrition from food image with a smartphone is one of those. It can aid users, especially the obese people, to measure and control their calorie intake personally.

Researchers for this task have proposed various methods. Some of them emphasized on the accuracy, some of them emphasized on the user flexibility. In this paper, we have proposed a novel method that can ensure more user flexibility as well as high accuracy. We propose capturing images from a known distance and avoid the presence of any reference object. Here, the user needs to capture an image of the food with a specific posture. With this, we calculate the distance of the camera from the food plate and ultimately measure the volume of the food using simple geometry. For solid food, calorie can be estimated using a single image from a specific posture. For liquid food, the distance between the camera and the liquid container has to be manually adjusted. We also address the case of transparent and nontransparent liquid container separately.

Although many studies used distance information to estimate the calorie of *solid* food, to the best of our knowledge, our work is the first approach using distance information to estimate the calorie content of *liquid* food in a *non-transparent container*. As the method is distance based, the overhead of carrying and segmenting the reference object is easily avoided. This, in turn, gives the system portability and user-flexibility under different circumstances. The proposed system for solid food estimates the distance from a well-specified posture of user's hand which makes estimation accuracy better and more reasonable compared to the other distance-based works such as [9]. Thus the combination of accuracy and flexibility makes this research unique in this field.

The organization of the paper is as follows. First, we describe some of the related works in Section II. We provide a brief overview of the proposed system in Section III and elaborate it in Section IV. Evaluation of our system is described in Section V. Section VI concludes the paper with pointers to some possible future works.

II. Related Works

Volume measurement from food image usually employs capturing images of food from suitable angles, segmentation of food portion from the image and taking suitable reference object for camera calibration. Different works addressed these sectors differently to measure food volume.

A fiducial marker like checkerboard has been employed for determining intrinsic and extrinsic parameters of a camera in many works [10]–[13]. These parameters are used to determine the actual measurement of the food item from the image. Also, other regular objects can be used instead of the checkerboard as a geometric reference for systems to work with smartphone's built-in cameras. For example, Kong et al. [14] use a credit card as a geometric reference object. Okamoto et al. [15] use credit card sized cards or wallet as a reference object. In all these scenarios the additional need for calibration card or the reference object placed beside the food item while taking an image creates inconvenience for the user. Also, this method requires additional computation for the segmentation of the reference object. Moreover, if that object is not differentiable from the background due to the similarity in texture, a manual segmentation might be required.

Some methods proposed avoiding the need for carrying the card. For example, Jia et al. used the food plate of known diameter as the geometrical reference [16]. However, it is not always applicable as users may eat food from restaurants where it is *near-impossible* to know the diameter of the plate. Even in residence, the user may not take food every time using the same plate. Pouladzadeh et al. proposed a method where they used the thumb of the user as a reference to measure the volume. Here, food image was captured placing user thumb beside the dish [17].

Though this method added some user flexibility by using their thumb as a reference, it has some problems. Firstly, the user's thumb size will vary with the different orientation of user's thumb or the camera which will contribute to error. Secondly, the method did not specify the target portion of the thumb for segmentation.

Kuhad et al. [9] proposed a distance-based volume measurement method where the user needs to capture photo from a known height which is assumed to be the height of the user. Then it captures the camera orientation information from the mobile sensor and uses it to calculate the real distance from the camera to the food. From this distance, the actual size of the food can be determined. The limitation of this approach is that it assumes the height of the food table plane from the ground is half of the user height which may not be true always. Our proposed method is also distance based and uses mobile sensor data to get camera orientation. In the proposed technique, we take a measurable distance that uses the length of the user's hand and the method does not depend on the height of the food table at all.

Calculation of the food volume also differs in various works. Chae et al. proposed food specific shape templates to have more accurate segmentation and gather geometric information which is difficult to achieve generally from a single image. The method used that information for 3D reconstruction of the food. They needed many templates to match varieties of foods to measure the volume accurately [11] .

Kong et al. [14] also employed volume estimation from 3D reconstruction from three images taken around the dish at approximately every 120° or from a video. Zhang et al. [12] used the virtual reality technique to estimate the volume. In this work, 3D virtual frames were created to fit food items. Though this type of methods could give higher accuracy, it has more computational costs than our method and in some cases requires more user interaction.

Okamoto et al. [15] proposed a method where calories are estimated directly from the food size estimated from the top view image. They used quadratic curve estimation from the 2D size of food following the assumption that the height of the food portion correlates with the size of the food and its category. We have employed a similar estimation under slightly different assumption. We assume a correlation between the volume and the area of food portion projected in the image.

III. Overview of the Proposed System

The proposed system avoids the need for maintaining and capturing any reference object as it creates much overhead from the users' perspective associated with carrying and correctly placing beside the food item. Also, there is an additional segmentation error attached to the reference object which in turn contributes to higher inaccuracy in estimating food volume. So, instead of using *reference object* based food volume estimation, we use *distance based* food volume estimation.

We employ different techniques based on the food type. For solid food items, we estimate the distance between food and camera from the *posture* of user's hand and device orientation. On the other hand, for liquid food items, image is captured from a predefined distance from the liquid container. Then both for solid and liquid, the distance information is used to find the centimeter to pixel ratio. We use color based K-means clustering for *segmentation* of food portion, and Convolutional Neural Network(CNN) for *recognition* of food items.

Fig. 1 shows the workflow of the proposed system. Different procedures need to be followed to process (i) solid, (ii) liquid in transparent containers, and (iii) liquid in nontransparent container. The user at first selects one of the three food types. Then based on the food type the system performs the successive steps as shown in fig. 1.

IV. Proposed System

In this section, we discuss the proposed system in detail. Image processing for solid and liquid food is different. So to enable the system to decide how to process the first image user needs to select between solid and liquid food options from the application menu.

A. Solid food in plate

In this scenario the centimeter to pixel ratios in horizontal and vertical direction are calculated first, then the different food regions are segmented, and various food regions are recognized. Finally, the volume and calorie information for each area are estimated. We illustrate each step here.

1) Estimate the distance between food item and camera: To calculate the centimeter to pixel ratio in horizontal and vertical direction, we estimate the distance between camera and plate and also the plate diameter from the image first. A user needs to capture that image in the following manner:

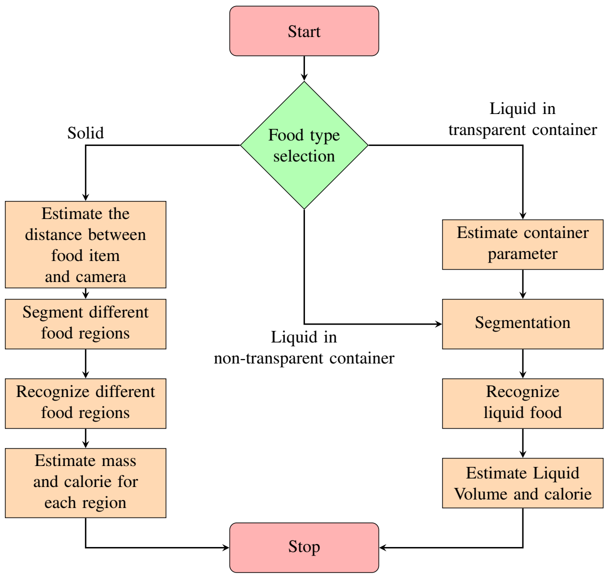

Fig. 1: System Flow Chart

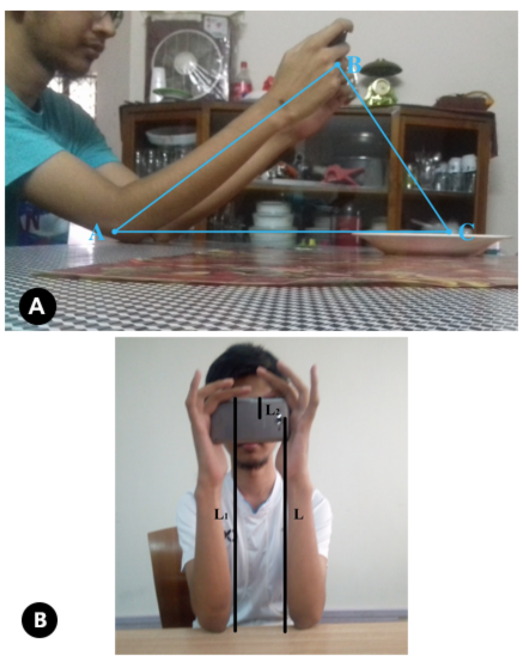

Fig. 2: Hand position A) Fixed posture B) Parallel position.

Let L_1 be the distance between user's elbow and the upper edge of smartphone(shown in Fig. 2 (B)). Let L_2 be the distance between the upper edge of the smartphone and the position of the built-in camera. L_1 and L_2 is registered when the application is first installed. So the distance between the user's elbow and the camera is $L = L_1 - L_2$.

The user must hold the smartphone with two index finger placing both elbows on the table. Right and left hand of the user should be parallel to each other as shown in Fig. 2 (B). Fig. 2 (A) shows the appropriate position of the hand while capturing the image. Ideally, the user's hand should form a

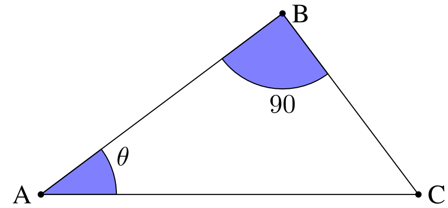

Fig. 3: Estimating distance between camera and food plate.

straight line perpendicular to the camera direction, and focus of the camera should be near the center of the plate.

Let AB represent the hand holding the camera. $AB = L$ (see Fig. 3). Point C is the intersection of the optical axis of the camera and the plate and is near the center of the plate. The $\angle ABC$ is 90 degrees, and $\angle BAC = \theta$ can be found from the tilt sensor of the device. The distance between plate and camera BC can be calculated as follows using simple trigonometry.

$$\frac{BC}{AB} = \tan(\angle BAC) \Leftrightarrow \frac{BC}{L} = \tan(\theta) \Leftrightarrow BC = L\tan(\theta)$$

The estimated distance between the camera and the food plate is then used to find out the diameter of the plate. The calculation procedure is as follows:

As camera direction is not perpendicular to the table plane, plate diameter in horizontal and vertical direction will always be different. Let P_h be the plate diameter observed in the horizontal direction. Let S be the distance of the plate from the camera. Let D be the actual diameter of the plate. Now the plate diameter perceived by the camera in the projection plane is proportional to the actual width of the plate and inversely proportional to the distance between the plate and the camera.

$$P_h \propto \frac{D}{S} \Leftrightarrow P_h = k\frac{D}{S}$$

Here k is a camera specific constant that depends on camera parameters. We estimate the value of k by placing an object of known size at a known distance from the camera before first use. We assume that when our system is working, the value of k for the current camera is already known. Now, the actual diameter of the plate is,

$$D = \frac{P_h \times S}{k} \tag{1}$$

Actual plate diameter D is then used to calculate centimeter to pixel ratio in horizontal and vertical direction.

Let P_v be the plate diameter observed in the vertical direction regarding pixels. Let r_h, and r_v be the centimeter to pixel ratio in a horizontal and vertical direction respectively. r_h and r_v can be determined as follows.

$$r_h = \frac{P_h}{D}, \text{ and } r_v = \frac{P_v}{D}$$

r_h and r_v has been used to estimate food portion area in square centimeters later on.

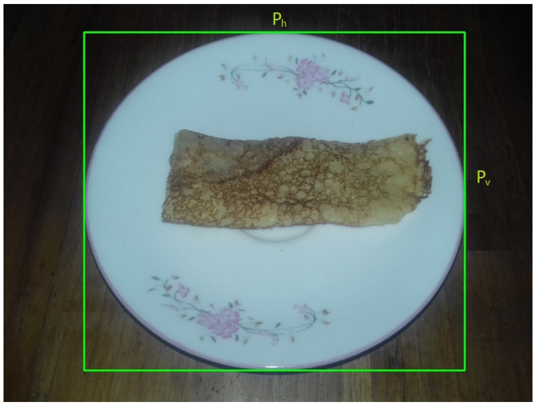

Fig. 4: Plate observed from captured image

2) Segment different food regions: At this step, we separate all different food items in the image and ignore all non-food items. Here we assume that background, food plate and the food do not share the same color. At first k-means clustering is applied on the image with the value $k = 3$. The background behind the plate, the plate itself and food items are expected to be in different clusters. After this, Otsu's thresholding method [18] is applied to the image. At last Canny edge detector [19] is applied. The steps are shown in Fig. 5.

We detect the plate location using an ellipse detector. Every region inside this plate is expected to be a food item(s). We identify each different region using contour detection. After this, four extreme points of each contour would be located first – left, right, top and bottom. A bounding box would be formed using these four points. Then using this bounding box the food region from the original image is segmented out. This image segment will be the input in the next step – the food recognition step. Notably, the system can support mixed food items because the calorie of each food portion is estimated separately.

3) Recognizing each food segment: For recognition of each food item, we use the Convolutional Neural Network(CNN). Our CNN model is based on the system proposed in [20]. We trained our system using FooDD dataset [21]. This dataset contains 3000 images of 30 food categories. The images in this dataset were taken from different cameras under different lighting conditions.

4) Estimating mass and calorie for each food segment: We try to estimate the mass of the food from the area of the 2D projection. We calculate the area of the contour representing the food portion. This contour area is the number of pixels occupied by the contour. We estimate the area regarding square centimeters by using centimeter to pixel ratio (r_h and r_v).

$$A = \frac{\text{Contour area in pixels}}{r_h \times r_v}$$

We convert the area to food amount using food specific conversion formula. In general, most food items do not have a fixed height. So it is not practical to convert the area into food mass using a linear equation. We assume that when the amount of food increases, length, width and height of the food portion increases uniformly. Thus we assume that the height of the food item is correlated to its area. Let W, A, V be the width, area, and volume of the food portion respectively. Then, $A \propto W^2$, and $V \propto W^3$. The relation between volume and projected area is, $V \propto A^{\frac{3}{2}}$. If the mass of given food is M and density of the food is constant, then we have

$$M \propto V \Leftrightarrow M \propto A^{\frac{3}{2}} \Leftrightarrow M = kA^{\frac{3}{2}} \tag{2}$$

k in the above equation is a food-specific constant. The value of k for each food is estimated as follows. First, we took samples of a food item and measured their mass. Then, we took pictures of these samples and determined the projected area of the food items. We estimated the value of k using curve fitting with the least squares method.

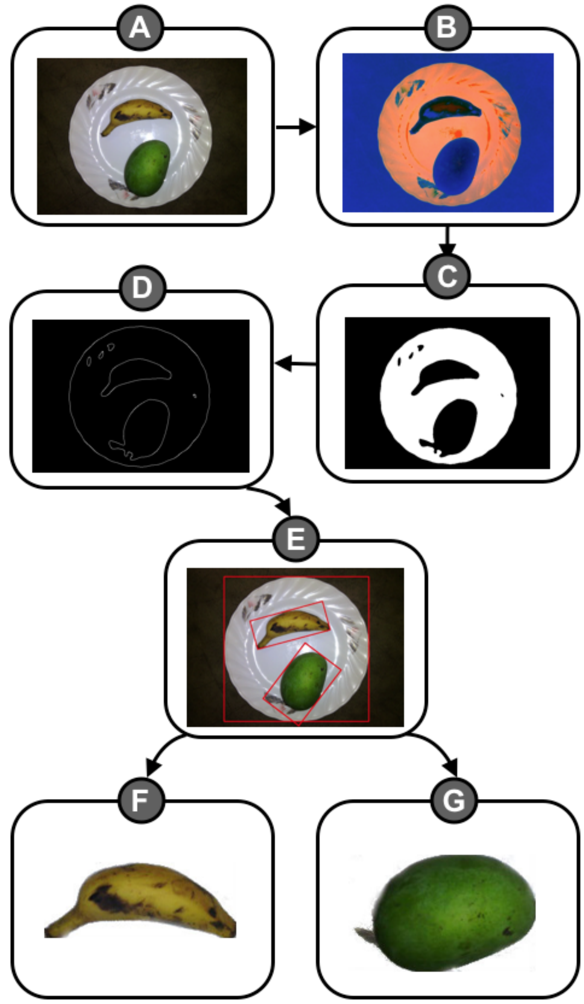

Fig. 5: Steps in separating each food portion A) Original Image B) K-means clustering C) Otsu thresholding D) Canny edge detector E) Detect bounding box F) Food portion 1 G) Food portion 2.

B. Liquid food in transparent container

We assume that the liquid food (milk, tea, etc.) is contained in a glass. The user has to capture the photo of the glass from

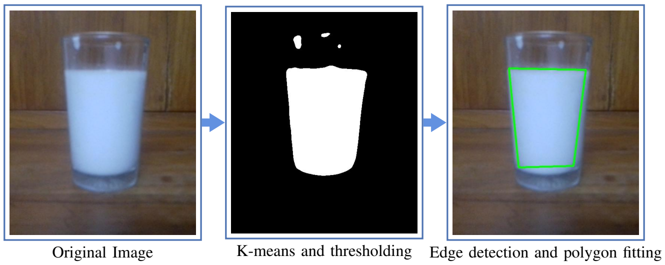

Fig. 6: Processing liquid image for transparent container

a known distance. This helps to find out the actual size and shape of the liquid. The user's hand length has been used as this known distance (see Fig. 7).

1) Segmentation of liquid: We segment out the liquid from the transparent container with along the background. First, we apply the smoothing filter to the image to reduce noise. Then we run k-means clustering on the image. After this, otsu's thresholding method [18] is applied. Then we use Canny edge detector operator [19] to segment out the liquid portion. Some of the contour formed after edge detection is due to the reflection in the glass and the background noise. But usually, these noises do not create large contours. So we filter out these by checking the size of the contours. The workflow of the system is shown in Fig. 6.

2) Recognition of liquid: Recognition step is similar to the solid food. Once the liquid portion is separated, it is sent to the trained classifier as input.

3) Estimation of volume and mass: Here we use the contour found from the segmentation step. Because of the roundness of the glass and effect of blurring, detected contour from edge detector may have some irregularity in its border. We fit this contour to a quadrilateral using polygon fitting. Dimensions of the liquid regarding pixels is estimated from aspects of the quadrilateral. Then these dimensions are converted to real-world dimensions using distance information as in equation 1. Let R, r and h be the upper radius, lower radius, and height of this liquid portion respectively.

The volume of the liquid can be calculated using the following formula [22].

$$V = \frac{\pi h}{3}(R^2 + Rr + r^2) \tag{3}$$

If the mass and density of the liquid is M and ρ respectively then we can estimate the mass as follows:

$$M = \rho V \tag{4}$$

C. Liquid food in nontransparent container

For nontransparent containers, we need to estimate the size and shape information of the container. We also need to know the height of the liquid in that container. It is challenging to get all these information from a single image. So we use two different images. The first image must be taken from a

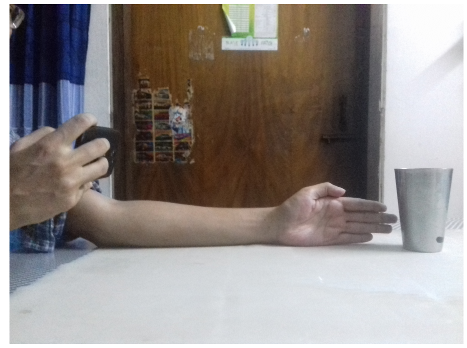

Fig. 7: Capturing image from fixed distance

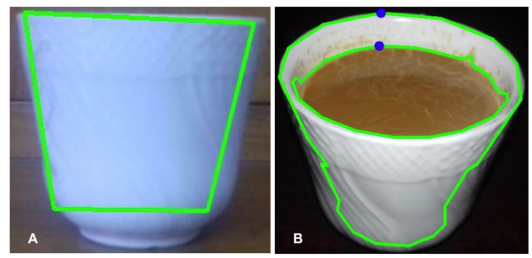

Fig. 8: Tea cup A) Container parameters B) Nontransparent container

known distance from the side of the container. The length of the user's hand is used as the known distance (see Fig. 7). This image is used for estimating the size and shape of the liquid container. The second image should be taken from an angle so that some portion of the liquid is visible. The user does not have to maintain a fixed distance for this image. This image is used to estimate the height of the liquid in the container. Detail steps are described next.

1) Estimating container parameters: We assume that the container has the shape of a cone frustum (see Fig. 8 (A)). So we estimate the upper radius, lower radius and the height of the container using first image and distance information.

2) Recognition: When the second image is taken, some portion of the liquid should be visible (Fig. 8 (B)). Liquid food is separated from the background using color-based K-means clustering and contour detection. This process is similar to the method used for solid food. This portion is sent to a trained classifier for recognition.

3) Volume and calorie estimation: After K-means clustering and edge detection, we expect to get two separate contours for the liquid container and the liquid food (see Fig. 8 (B)). The closed region formed by the liquid is inside the closed area formed by the container. We detect the first point of both contours and use them to estimate the liquid height inside the container.

We use the difference between two first points to estimate the height of liquid. Fig. 9 shows the geometry of the estimation. Let $ABCD$ represent the container, and YZ represent the liquid height inside the container. AF and YG are rays reaching the camera. We assume that AY is negligible compared to the distance between the camera and the cup.

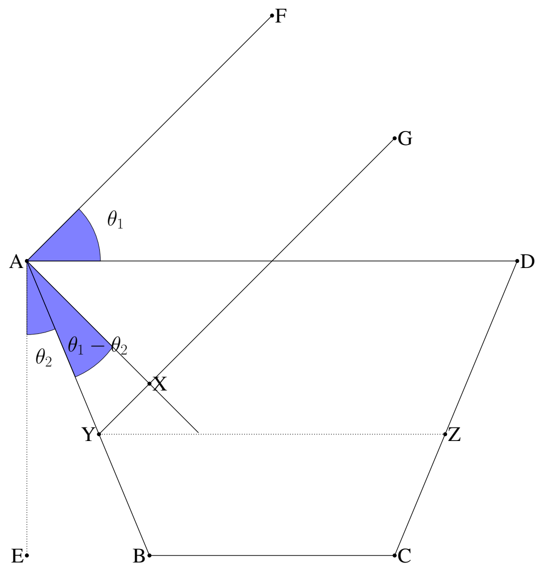

Fig. 9: Estimating liquid height inside container

Thus, AF and YG are almost parallel. We need to estimate the length AY. But the length AY may not be parallel to the projection plane. So we imagine a plane AX parallel to the projection plane. The length AY would appear to be length AX because of this angle. AX can be estimated by comparing it with the horizontal diameter of the container. The width of the container is already derived from the first image.

Let $\angle FAD = \theta_1$ and $\angle EAB = \theta_2$. θ_1 is calculated from the horizontal and vertical diameters of the cup.

$$\theta_1 = \sin^{-1}\frac{b}{a}$$

where b is the vertical diameter and a is the horizontal diameter of the cup both perceived in the camera. θ_2 is calculated from the first image. In $\triangle AXY$ we have,

$$\frac{AX}{AY} = \cos\angle XAY$$
$$\frac{AX}{AY} = \cos(\theta_1-\theta_2)$$
$$AY = \frac{AX}{\cos(\theta_1-\theta_2)}$$

AX, θ_1 and θ_2 is already known so the equation above is used to calculate AY. Now properties of the similar triangle are used to calculate the upper radius and height of liquid portion. After that volume of the liquid is estimated using Equation 3. Equation 4 is then used to convert the volume to liquid mass.

V. Evaluation

For food recognition, our CNN model achieved 96% accuracy with five-fold cross-validation on the FooDD dataset [21].

To calculate calorie estimation accuracy, we randomly chose eight food items including two liquid food items. To determine the parameter k of the equation 2, we measured the actual mass (M) of the food items using a weight meter. Then we estimated the projected area (A) of the food items from the captured images. For each food item, we plotted the mass (M) against the projected area (A) and then used curve fitting with *least square method* to estimate the value of parameter k.

We ran *leave-one-out* cross-validation with the captured images to get the accuracy of the estimation. We assume that calorie in a unit mass of food is constant. Calorie in food is calculated from the mass with the help of USDA National Nutrient Database and other Internet resources [23], [24].

For liquid food items, we estimated the upper radius, lower radius, and height of the liquid portion from the image. Then the volume of liquid was estimated using Equation 3. The actual volume of liquid was determined using a volumetric flask. Both actual and estimated volume was converted to mass and calorie was estimated using nutrition database [23], [24]. The error was calculated by comparing the calorie from real volume and estimated volume.

To determine the accuracy of the proposed system, we used two performance metrics:–(i) Standard Error (SE), and (ii) Relative Standard Error (RSE). They were calculated as follows:

Standard Error (SE) Calculation. Suppose Y is the calorie calculated from the actual food mass, Y' is the calorie estimated from the image, and N is the number of observations. Then, the Standard Error, SE will be:

$$SE = \sqrt{\frac{\sum(Y-Y')^2}{N}}$$

Relative Standard Error (RSE) Calculation. Suppose the average calorie of a food item from N observations is denoted by $\bar{Y}$. Then RSE will be:

$$RSE = \frac{SE \times 100}{\bar{Y}}$$

The standard error (SE) reflects the error of calorie estimation in serving size, and the relative standard error (RSE) demonstrates the failure in every 100 calories. The SE and RSE for both solid and liquid food items are shown in Table I.

TABLE I: SE and RSE of eight food items

Food name	Cal./gm	K	SE	RSE
Rice	3.65	0.076	9.02	1.2%
Apple	0.52	0.765	2.83	4.01%
Orange	0.97	0.85	8.63	4.657%
Banana	0.89	0.36	0.78	1.5%
Mango	0.60	0.495	12.75	6.48%
Patisapta	1.05	0.956	17.27	8.11%
Milk	0.63	N/A	5.53	4.08%
Tea*	0.18	N/A	2.88	5.93%

*+1 spoonful of sugar, SE= Standard Error (in calory)
K=Estimated parameter, RSE=Relative Standard Error

Patisapta has the highest relative standard error (RSE) of 8.11%. The reason for high RSE value is that the volume of this food can vary largely due to its serving method. An example image of patisapta can be seen in Fig. 4. Because of the varying presentation style and serving size, it is difficult to achieve very high accuracy for this food. Error in case of fruits like apple, mango, etc. is largely due to their irregular shape and non-uniform density. Average standard error for all food items is 7.46 calories, and the average relative standard error for all food items is 4.46%. This result is reliable given the fact that the portability and flexibility of our system are very high.

VI. Conclusion and future work

In this paper, we propose an automated calorie estimation system based on the distance between the camera and the target. The system can handle multiple food items on the same plate and applies separate methods for handling liquid food in the transparent and non-transparent container. Experimental results show promising accuracy with greater user flexibility and portability of the simple system that we propose. The major sources of error in calorie content estimation are mainly due to different serving style, irregular shape combined with the non-uniform density of some solid natural food items.

In the future, we want to improve our recognition system so that it can recognize more food items. We also want to use more advanced image segmentation method like grab cut [25] that allows more accuracy based on user interaction.

Acknowledgement

The research grant supports this research under the Higher Education Quality Enhancement Project (HEQEP) CP-3137 awarded by World Bank and approved by University Grant Commission (UGC), Bangladesh.

References

[1] Obesity and overweight. (n.d.). http://www.who.int/mediacentre/factsheets/fs311/en/. Accessed: 2017-02-16.

[2] Md Kamrul Hasan, Munirul Haque, Nazmus Sakib, Richard Love, and Sheikh I Ahamed. Smartphone-based human hemoglobin level measurement analyzing pixel intensity of a fingertip video on different color spaces. *Smart Health*, 5:26–39, 2018.

[3] Md Kamrul Hasan, Nazmus Sakib, Joshua Field, Richard R Love, and Sheikh I Ahamed. A novel technique of noninvasive hemoglobin level measurement using hsv value of fingertip image. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, pages 221–229. IEEE, 2017.

[4] Md Kamrul Hasan, Nazmus Sakib, Richard R Love, and Sheikh I Ahamed. Rgb pixel analysis of fingertip video image captured from sickle cell patient with low and high level of hemoglobin. In *Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017 IEEE 8th Annual*, pages 499–505. IEEE, 2017.

[5] Md Kamrul Hasan, Nazmus Sakib, Richard R Love, and Sheikh I Ahamed. Analyzing the existing noninvasive hemoglobin measurement techniques. In *Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017 IEEE 8th Annual*, pages 442–448. IEEE, 2017.

[6] Golam MT Ahsan, Md O Gani, Md K Hasan, Sheikh I Ahamed, William Chu, Mohammad Adibuzzaman, and Joshua Field. A novel real-time non-invasive hemoglobin level detection using video images from smartphone camera. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, pages 967–972. IEEE, 2017.

[7] Md Kamrul Hasan, Golam Mushih Tanimul Ahsan, Sheikh Iqbal Ahamed, Rechard Love, and Reza Salim. Pain level detection from facial image captured by smartphone. *Journal of Information Processing*, 24(4):598–608, 2016.

[8] Ferdaus Kawsar, Md Kamrul Hasan, Tanvir Roushan, Sheikh Iqbal Ahamed, William C Chu, and Richard Love. Activity detection using time-delay embedding in multi-modal sensor system. In *International Conference on Smart Homes and Health Telematics*, pages 489–499. Springer, 2016.

[9] Pallavi Kuhad, Abdulsalam Yassine, and Shervin Shimohammadi. Using distance estimation and deep learning to simplify calibration in food calorie measurement. In *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2015 IEEE International Conference on*, pages 1–6. IEEE, 2015.

[10] Mingui Sun, Qiang Liu, Karel Schmidt, Jie Yang, Ning Yao, John D Fernstrom, Madelyn H Fernstrom, James P DeLany, and Robert J Sclabassi. Determination of food portion size by image processing. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 871–874. IEEE, 2008.

[11] Junghoon Chae, Insoo Woo, SungYe Kim, Ross Maciejewski, Fengqing Zhu, Edward J Delp, Carol J Boushey, and David S Ebert. Volume estimation using food specific shape templates in mobile image-based dietary assessment. In *Computational Imaging*, page 78730K, 2011.

[12] Zhengnan Zhang, Yongquan Yang, Yaofeng Yue, John D Fernstrom, Wenyan Jia, and Mingui Sun. Food volume estimation from a single image using virtual reality technology. In *Bioengineering Conference (NEBEC), 2011 IEEE 37th Annual Northeast*, pages 1–2. IEEE, 2011.

[13] Mei-Yun Chen, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung. Automatic chinese food identification and quantity estimation. In *SIGGRAPH Asia 2012 Technical Briefs*, page 29. ACM, 2012.

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

[15] Koichi Okamoto and Keiji Yanai. An automatic calorie estimation system of food images on a smartphone. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pages 63–70. ACM, 2016.

[16] Wenyan Jia, Yaofeng Yue, John D Fernstrom, Zhengnan Zhang, Yongquan Yang, and Mingui Sun. 3d localization of circular feature in 2d image and application to food volume estimation. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4545–4548. IEEE, 2012.

[17] Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement*, 63(8):1947–1956, 2014.

[18] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[19] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[20] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. Food image recognition using very deep convolutional networks. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pages 41–49. ACM, 2016.

[21] Parisa Pouladzadeh, Abdulsalam Yassine, and Shervin Shirmohammadi. Foodd: food detection dataset for calorie measurement using food images. In *International Conference on Image Analysis and Processing*, pages 441–448. Springer, 2015.

[22] Jim wilson, the university of georgia. formulas. http://jwilson.coe.uga.edu/emt725/Frustum/Formulas.html. Accessed: 2017-07-26.

[23] Calorie control council. http://caloriecontrol.org/healthy-weight-tool-kit/food-calorie-calculator/? Accessed: 2017-08-15.

[24] Calories in my homemade patishapta. http://www.myfitnesspal.com/food/calories/my-homemade-patishapta-55351933. Accessed: 2017-08-15.

[25] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.