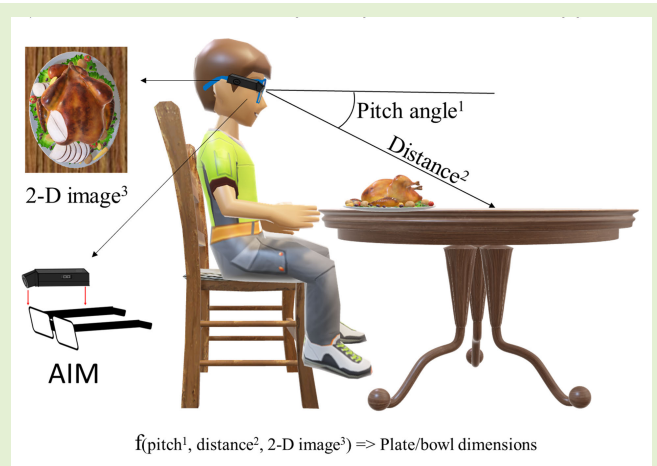


# Estimation of Plate and Bowl Dimensions for Food Portion Size Assessment in a Wearable Sensor System

Viprav B. Raju<sup>1</sup>, Student Member, IEEE, Delwar Hossain<sup>1</sup>, Student Member, IEEE, and Edward Sazonov<sup>1</sup>, Senior Member, IEEE

**Abstract**—Automatic food portion size estimation (FPSE) with minimal user burden is a challenging task. Most of the existing FPSE methods use fiducial markers and/or virtual models as dimensional references. An alternative approach is to estimate the dimensions of the eating containers prior to estimating the portion size. In this article, we propose a wearable sensor system (the automatic ingestion monitor integrated with a ranging sensor) and a related method for the estimation of dimensions of plates and bowls. The contributions of this study are: 1) the model eliminates the need for fiducial markers; 2) the camera system [automatic ingestion monitor version 2 (AIM-2)] is not restricted in terms of positioning relative to the food item; 3) our model accounts for radial lens distortion caused due to lens aberrations; 4) a ranging sensor directly gives the distance between the sensor and the eating surface; 5) the model is not restricted to circular plates; and 6) the proposed system implements a passive method that can be used for assessment of container dimensions with minimum user interaction. The error rates (mean  $\pm$  std. dev) for dimension estimation were  $2.01\% \pm 4.10\%$  for plate widths/diameters,  $2.75\% \pm 38.11\%$  for bowl heights, and  $4.58\% \pm 6.78\%$  for bowl diameters.

**Index Terms**—Dietary assessment, food imaging, food portion, food volume, portion size estimation, wearable sensors, wearable technology.



## I. INTRODUCTION

RELIABLE and accurate portion size estimation is challenging but essential for dietary assessment. Image-based dietary assessment has been one of the fastest growing areas of research in this milieu. Image-based assessment can be split into manual assessment and automatic assessment. Manual assessment can be done using digital food records [2] or by image-based 24-h recall/self-reporting that

involve food atlases [3], [4], [5], [6], [7]. The image-based food records involve capturing meal images that are reviewed later by the participant or by a professional (nutritionist/clinician/researcher) to estimate the portion size. Digital food images are a useful tool for the quantification of food items and in portion size estimation [8], [9], [10]. Images of food leftovers are also captured in some studies, which improved the portion size accuracy [11]. Recall or self-reporting methods use food atlases. Food atlases are reference guides that are taken to present various portions representative of the range of portion sizes usually consumed. Either during or after data collection, participants are asked to report the food quantity consumed by selecting a particular image, a fraction/multiple of an image, or a combination of images [12].

The abovementioned manual methods are cumbersome, subject to memory (and therefore prone to error), and are not accurate compared to the much recent automatic assessment methods. A previous review [13] identified some of the existing image-based food portion size estimation (FPSE)

Manuscript received 10 November 2022; revised 5 January 2023; accepted 7 January 2023. Date of publication 24 January 2023; date of current version 28 February 2023. This work was supported in part by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, under Award R01DK100796 and Award R01ADK122473. The associate editor coordinating the review of this article and approving it for publication was Prof. Subhas C. Mukhopadhyay. (Corresponding author: Viprav B. Raju.)

The authors are with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35401 USA (e-mail: vbraju@crimson.ua.edu; dhossain@crimson.ua.edu; esazonov@eng.ua.edu).

Digital Object Identifier 10.1109/JSEN.2023.3235956

methods that are automatic. It was seen that food portion size can be estimated automatically using food images captured during the meal [14], [15], [16], [17], [18]. However, automatic FPSE from food images is a challenging task since the two-dimensional (2-D) image lacks the three-dimensional (3-D) real-world information. There is a lack of reference to measure/gauge the size/volume of the food items. To tackle this problem, the dimensional reference is obtained by using a visual cue that must be present in a food picture. A few methods used virtual objects or objects that already exist in a typical food image to aid in FPSE. Some of the popular approaches included geometric models [19], VR-based referencing [20], circular object referencing [17], [21], and thumb-based referencing [22]. Shang et al. [23] used a structured light-based 3-D reconstruction approach to estimate food volume. Jia et al. [17] used the “plate method” for FPSE where the circular plates present in the image are used to determine location, orientation, and volume of the food items. The study, however, only considers circular plates.

A fiducial marker of known dimensions placed in the images can also be used as a point of reference [17], [22], [24], [25]. The type of reference used determines the complexity of setup. Some methods require the users to carry around the reference (checkerboards, blocks, and cards) and some require special dining setups, which increases user burden.

Another classification in image-based FPSE can be done based on the mode of image capture. Food image capture can either be *active* or *passive*. Active methods rely on the participant to capture the food image by a camera (such as a smartphone camera), typically, before and after an eating episode. The images are then processed using computer vision models to segment foods, recognize foods, estimate portion size/volume, and compute energy content [26], [27], [28]. Active methods provide detailed information such as meal timing, location, and duration of the eating episodes. However, these methods require the active participation of the users, which can be a burden. Some of the active methods that predict portion sizes require fiducial markers in the food image to assist manual review/computer algorithms [26], [29]. The placement of these markers combined with the active nature of image capture increases the user burden considerably.

One study presented a new *active* method for food volume estimation without using a fiducial marker. The method utilizes a special picture-taking strategy on a smartphone [1]. A mathematical model that uses the height and orientation of the smartphone was used to determine the real-world coordinates of the plane of the eating surface in the capture image. Bucher et al. [30] presented and tested a new virtual reality method for food volume estimation using the International Food Unit. This method, however, requires the user to place the smartphone on the eating surface while image capture and also needs additional user interaction in using the virtual International Food Unit.

Food images can also be acquired by a “passive” method using wearable devices that capture images continuously (both food and nonfood) without the active participation of the user [31], [32]. The passive methods minimize the burden of active capture using a wearable camera. However, FPSE meth-

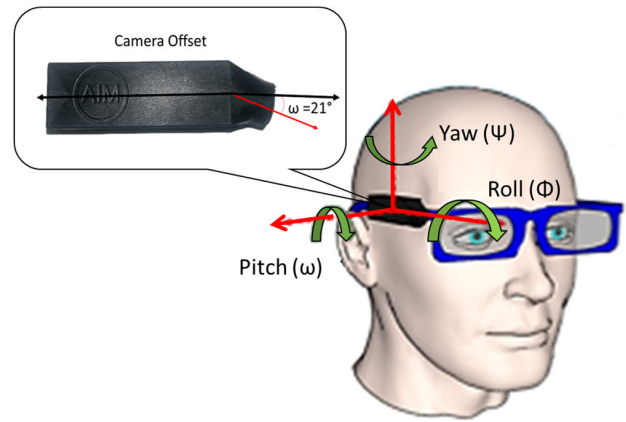


Fig. 1. Rotational axes of the AIM sensor. Also depicted is the camera offset of  $21^\circ$  with respect to the axis of the eyeglass/AIM device.

ods that require fiducial markers cannot be easily integrated with the passive image capture since the user is not actively taking images and do not know when to place these markers.

The automatic ingestion monitor [33] is a wearable sensor system [automatic ingestion monitor version 2 (AIM-2)] that is mounted on a user’s eyeglass. The sensor consists of a combination of sensors for accurate detection of food intake and triggering of a wearable camera (*passive*). In this study, we integrate a time-of-flight (ToF) ranging sensor with AIM-2 and propose a novel method to determine container dimensions (bowls/plates). The method does not require fiducial markers. Once the size of the vessels is known, portion size can be estimated using the “plate method.” In this study, our objective is to measure the dimensions of plates and bowls.

The major contributions of the proposed work are: 1) the model eliminates the need for fiducial markers; 2) the camera system (AIM-2) is not restricted in terms of positioning relative to the food item; 3) our model accounts for radial lens distortion in caused due to lens aberrations; 4) a ranging (ToF) sensor directly gives the distance between the sensor and the eating surface; 5) the model is not restricted to circular plates; and 6) a passive method that can be used for assessment of container dimensions with minimum user interaction.

## II. METHODS

### A. Equipment

In this study, a novel wearable sensor system (AIM-2 with a ToF ranging sensor) was used [33]. AIM-2 consists of a sensor module, which houses a miniature 5-Mpixel camera with  $120^\circ$  wide-angle gaze-aligned lens, a low-power 3-D accelerometer (ADXL362 from Analog Devices, Norwood, MA, USA), and a ToF ranging sensor (VL53L0X from STMicroelectronics). The sensor system is enclosed in a custom-designed 3-D printed enclosure. The ToF sensor is aligned with the camera axis.

The camera continuously captured images at a rate of one image per 10-s interval continuously throughout the day. Data from the accelerometer and ToF sensor were sampled at 128 Hz. All collected sensor signals and captured images were stored on an SD card and processed off-line in MATLAB (MathWorks Inc., Natick, MA, USA) for algorithm development and validation. The AIM enclosure is such that the

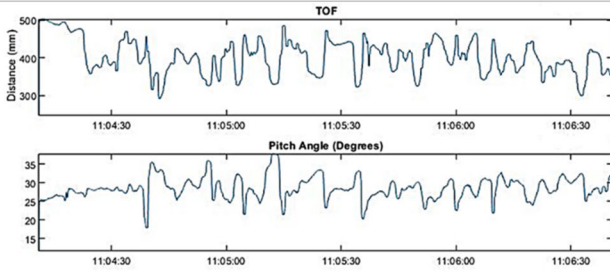


Fig. 2. ToF distance and pitch angle as a function of time (y-axis: hh:mm:ss) for a sample meal where the AIM-2 was used.

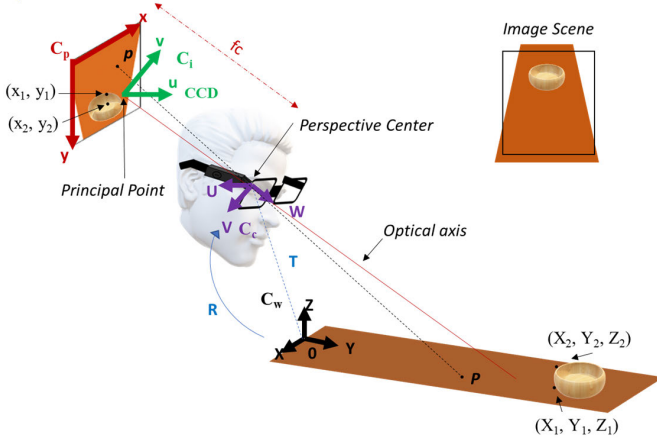


Fig. 3. Model of the AIM imaging system.

camera and the ToF sensor are at an angle of  $21^\circ$  with respect to the accelerometer axis, as shown in Fig. 1. We will be using this offset ( $+21^\circ$ ) while calculating the pitch of the camera. The raw sensor data from the accelerometer were preprocessed before extracting the pitch angle. A high-pass filter with a cutoff frequency of 0.1 Hz was applied to remove the dc component from the accelerometer signal.

The sensor pitch was calculated as in [33] and the device pitch is obtained by adding the offset ( $21^\circ$ ) to the sensor pitch. The distance readings are more straightforward, the raw values depicting the actual distances. Next, using the timestamp of an image, the respective pitch and distance readings were extracted. Fig. 2 shows the ToF distance readings and pitch plotted as a function of time for a sample meal.

## B. Geometric Model

The objective is to project the points in an image in the real-world coordinates. In this study, our primary concern is to measure the dimensions of the plate and bowls.

Refer to Fig. 3. Let  $P$  be a point in the world,  $C_w$  be a world coordinate system, and  $(XYZ)^t$  be the coordinates of  $P$  in  $C_w$ . Define the camera coordinate system,  $C_c$ , to have its  $W$ -axis parallel with the optical axis of the camera lens, its  $U$ -axis parallel with the  $u$ -axis of  $C_i$  (image coordinate plane), and origin located at the perspective center. Let  $(UVW)^t$  be the coordinates of  $P$  in  $C_c$ . The coordinates  $(UVW)^t$  are related

to  $(XYZ)^t$  by a rigid body coordinate transformation

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + T \quad (1)$$

where  $R$  is a  $3 \times 3$  rotation matrix and  $T$  is a  $3 \times 1$  translation vector.  $R$  is dependent on three angles of rotation, namely, pitch ( $\omega$ ), roll ( $\Phi$ ), and yaw ( $\Psi$ ). The three angles for the AIM device are shown in Fig. 1.

The principal point is the intersection of the imaging plane with the optical axis. Let  $f_c$  be the focal length of the lens of the imaging system. Define the 2-D image coordinate system,  $C_i$ , to be in the image plane with its origin located at the principal point,  $u$ -axis in the fast scan direction (horizontal rows of pixels on the sensor), and  $v$ -axis in the slow scan direction (vertical rows of pixels on the sensor) direction of the camera sensor. Fast scan indicates the pixel direction in which the sensor scans at a higher rate. Let  $p$  be the projection of  $P$  onto the image plane and let  $(\bar{u}, \bar{v})^t$  be the coordinates of  $p$  in  $C_i$ . The focal length (Table. I) of  $(\bar{u}, \bar{v})^t$  is given by

$$\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} = \frac{f_c}{W} \begin{pmatrix} U \\ V \end{pmatrix}. \quad (2)$$

Next, radial lens distortion is incorporated into the model in the following way. Let  $(uv)^t$  be the actual observed image point after being subject to lens distortion. Then,  $(u v)^t$  is related to  $(\bar{u} \bar{v})^t$  by

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} \left[ 1 + K_c (\bar{u}^2 + \bar{v}^2) \right] \quad (3)$$

where  $K_c$  is a coefficient, which controls the amount of radial distortion.

Finally, it is necessary to model the image sampling performed by the camera sensor [charged coupled device (CCD)]. A camera sensor consists of a 2-D array of photosensors. Each photosensor converts incoming light into a digital signal by means of an analog-to-digital converter. To obtain color information, one "sensor pixel" is divided into a grid of photosensors, and different color filters are placed in front of these multiple photosensors. Each of these photosensors receives light through only one of the three filters: blue, red, and green. Combining these measurements gives one color triple: (red intensity, green intensity, and blue intensity). This is known as the Bayer filter. Therefore, the digital image coordinates are not the same as the pixel coordinates.

Let  $C_p$  be the pixel coordinate system associated with the digital image. The pixel coordinates are related to the image coordinates by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_c^x & K_c \\ 0 & s_c^y \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} c_c^x \\ c_c^y \end{pmatrix} \quad (4)$$

where  $s_c^x$  and  $s_c^y$  are scale factors (pixel/mm),  $c_c^x$  and  $c_c^y$  are the pixel coordinates of the principal point, and  $K_c$  is the distortion coefficient (pixel/mm)

$$\begin{pmatrix} x \\ y \end{pmatrix} = f(XYZ). \quad (5)$$

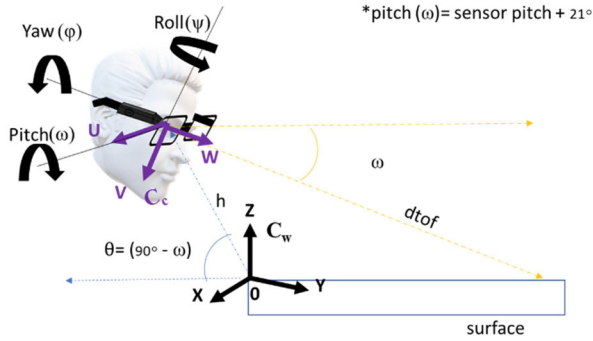


Fig. 4. Projection of camera coordinates on to the world coordinates. Also, depict the calculation of pitch angle of the camera.

We are interested in the inverse function of (5) for the purposes of dimension estimation, i.e.,

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = f^{-1}(xy). \quad (6)$$

With the known AIM sensor orientation, namely, the pitch angle of the sensor, provided by the inertial measurement unit, a right angle between the surface of the lens and the optical axis, and the projection relationship in (5), it can be shown that the inverse of function  $f$  in (6) exists for the tabletop, i.e.,

$$\begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} = f^{-1}(xy). \quad (7)$$

Note that  $Z = 0$  in (7) represents the plane equation of the tabletop.

Also, we assume that the roll ( $\Phi$ ) and yaw ( $\Psi$ ) of the sensor are zero. Therefore, the rotation matrix  $R$  is given by

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & -\sin\omega \\ 0 & \sin\omega & \cos\omega \end{bmatrix}. \quad (8)$$

From (4), the world coordinates of the tabletop are related to the pixel coordinates by

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1/s_c^x & 0 \\ -Kc & 1/s_c^y \end{pmatrix} + \begin{pmatrix} x - c_c^x \\ y - c_c^y \end{pmatrix}. \quad (9)$$

To calculate the translational and rotation matrices, we make use of the sensor pitch ( $\omega$ ) and distance readings from the AIM. The distance is obtained by the ToF sensor, and the sensor pitch is obtained by the accelerometer on the AIM device, as shown in Fig. 4. The camera on the AIM has an offset of  $21^\circ$

$$h = d_{tof} \times \tan(\omega) \quad (10)$$

where  $d_{tof}$  is the distance from the ToF sensor, which is the distance between the AIM and the eating surface. As in [1], we obtain the following equation:

$$W = \frac{h \cdot \sin(\theta)}{\cos(\theta) - \frac{v \cdot \sin(\theta)}{f_c}}. \quad (11)$$

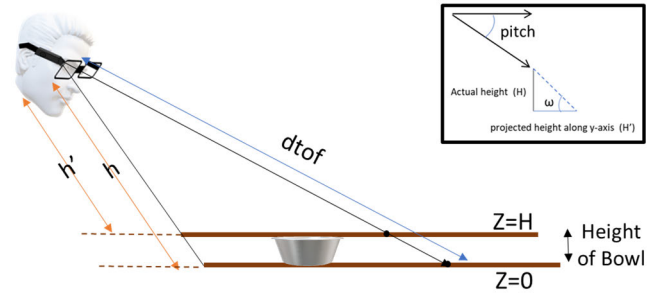


Fig. 5. Projection  $H'$  of height  $H$  on the  $y$ -axis. Calculating  $h'$  for obtaining the plane equation for  $Z = H$ .



Fig. 6. Left: test bench with the AIM attached to the tripod. Right: protractor used to measure the pitch angle of the sensor system.



Fig. 7. Test dataset: plates and bowls of varying sizes.

From (9), we obtain

$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{W}{f_c} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (12)$$

Finally, we obtain the equation

$$\begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} = R^{-1} \left( \begin{pmatrix} U \\ V \\ W \end{pmatrix} - T \right) \quad (13)$$

where  $T = [0; -h; 0]$ .

Equation (13) gives us the plane of the eating surface ( $Z = 0$ ).

The study mainly focuses on obtaining the dimensions of two types of vessels, namely, plates and bowls. We assume plates to be flat and part of the plane  $Z = 0$ . The heights/depths of the plates are assumed to be negligible and approximated to zero. We measure the dimensions of the plate on this plane.

However, in the case of bowls, first, the height of the bowl is measured along the  $y$ -axis, as shown in Fig. 5. The height is

TABLE I  
PARAMETERS OF THE AIM CAMERA

Focal Length ( $f_c$ )	2.4 mm
Sensor Size	3.67 mm x 2.74 mm
Image size (raw)	2592 x 1944
Image size (distortion corrected)	3378 x 2347
$c_c^x$	3378/2
$c_c^y$	2347/2
$s_c^x$	3378/3.67
$s_c^y$	2347/2.74

just a projection on the  $y$ -axis and the true height is calculated as in (14). Here, the assumption is that the bowl sides are flat and not curved

$$H = \tan(\omega) \times H'. \quad (14)$$

Once the height of the bowl ( $H$ ) is calculated, the equation of the plane  $Z = H$  is obtained instead of  $Z = 0$ . Fig. 5 shows the changes in the parameters for obtaining the adjusted plane equation

$$h' = h - (H \times \sec(\omega)) \quad (15)$$

where  $h$  is calculated as in (10).

Once  $h'$  is obtained, this value is plugged into (11) followed by (12) and (13).

We then measure the dimensions of the mouth of the bowls similar to the dimensions of the plates. We assume bowl of the mouth to be a part of the plane  $Z = H$ . The radius of the mouth of the bowl is then measured on this plane along the  $x$ - and  $y$ -axes.

### C. Data

The AIM sensor system was mounted on a test bench to collect data. The test bench consisted of a tripod and a protractor for angle measurement (see Fig. 6). The AIM device was placed on a tripod in front of a table at three pitch angles ( $40^\circ$ ,  $55^\circ$ , and  $70^\circ$ ) with respect to the parallel to the ground, at three different heights from the table surface (20, 35, and 50 cm). The angles were measured using the protractor fixed to the side of the sensor aligned with the camera (as shown in Fig. 6). The protractor was also calibrated to test for errors in the pitch angle measurement. The calibration was done in increments of  $10^\circ$  from  $0^\circ$  to  $70^\circ$ . The error in measurement was (mean  $\pm$  std. dev)  $-2.43^\circ \pm 1.36^\circ$ . The roll and yaw of the cameras were approximately set to be 0 for experimentation. Also, the roll and yaw for the AIM are assumed to be 0 when a person is eating.

Nine sets of data collected at a combination of three heights and three pitch angles were used for testing (see Fig. 7). A set of eight objects, three circular plates (diameter: small 18 cm, medium 22 cm, and large 26 cm), two square plates (side: small 18 cm and medium 23 cm), and three circular bowls, were used as objects of interest.

As a final step, four research assistants used the proposed methodology to estimate the bowl/container sizes of 3 (two circular bowls and one hollow rectangular box) shown in Fig. 8. The AIM device was worn by a user and a minimum of



Fig. 8. Validation images in a real-case scenario where AIM-2 was worn by a user on the eyeglass.

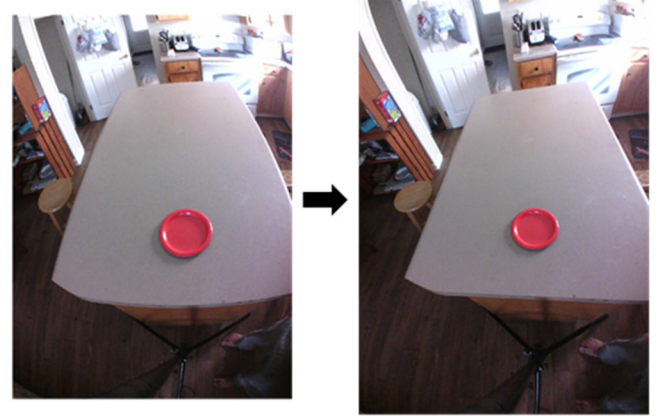


Fig. 9. Correcting distortion due to lens aberrations [see (3)].

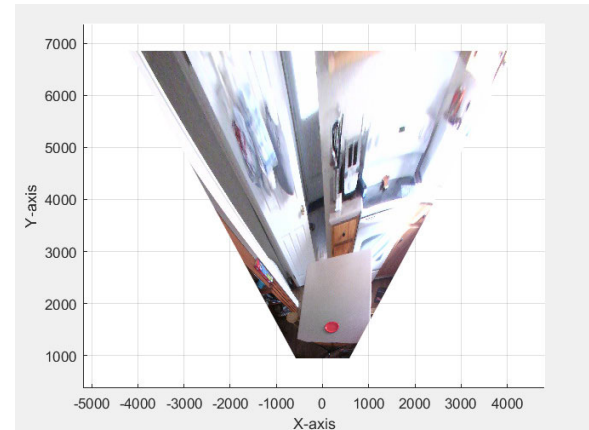


Fig. 10. Projection of the image on the plane equation  $Z = 0$  (all units are in mm).

three images were taken for each case without any restrictions on the position/tilt of the head.

The images and the sensor signals captured by the AIM at each setup were extracted and used as input to the model. The ground-truth dimensions were measured using a tape measure.

## III. RESULTS

Fig. 9 represents a sample result of the lens corrections after (3).

Using the world coordinates of the plane  $Z = 0$  and the projected image on the plane (see Fig. 10), the dimensions of plates were measured (see Fig. 11). Any object belonging to this plane can be measured using this projection.

Table II presents the results for the dimension estimation of plates using the proposed model. The error percentage in the dimension estimation of plates was (mean  $\pm$  std. dev)  $2.01\% \pm 4.10\%$ .

TABLE II  
ERROR IN DIMENSIONS FOR PLATES

Object	Height (cm)	Pitch (degrees)	Dtof (mm)	Predicted (cm)	Original (cm)	Error %	Mean Error (Mean $\pm$ Std.D)
Circular Plate (Small)	20	42	460	16.95	18	-5.84	-1.31 $\pm$ 4.82
		56	425	16.56	18	-8.00	
		64	390	19.07	18	5.94	
	35	38	740	17.13	18	-4.86	
		55	580	17.37	18	-3.48	
		72	550	18.92	18	5.11	
	50	41	870	17.61	18	-2.15	
		53	755	18.23	18	1.28	
		67	680	18.04	18	0.24	
Circular Plate (Medium)	20	42	460	20.59	22	-6.42	0.96 $\pm$ 4.66
		56	425	21.14	22	-3.93	
		64	390	23.36	22	6.17	
	35	38	740	22.51	22	2.30	
		55	580	21.47	22	-2.40	
		72	550	23.73	22	7.84	
	50	41	870	22.01	22	0.04	
		53	755	22.79	22	3.60	
		67	680	22.33	22	1.50	
Circular Plate (Large)	20	42	460	25.71	26	-1.12	5.28 $\pm$ 4.39
		56	425	25.91	26	-0.33	
		64	390	28.41	26	9.27	
	35	38	740	27.81	26	6.95	
		55	580	26.99	26	3.81	
		72	550	29.39	26	13.05	
	50	41	870	27.21	26	4.65	
		53	755	27.53	26	5.89	
		67	680	27.39	26	5.34	
Square Plate (Small)	20	42	460	17.31	18	-3.83	2.71 $\pm$ 4.38
		56	425	17.86	18	-0.80	
		64	390	19.37	18	7.62	
	35	38	740	18.33	18	1.81	
		55	580	18.34	18	1.91	
		72	550	19.88	18	10.46	
	50	41	870	18.92	18	5.10	
		53	755	18.03	18	0.18	
		67	680	18.35	18	1.93	
Square Plate (Medium)	20	42	460	21.98	23	-4.43	2.40 $\pm$ 4.73
		56	425	22.14	23	-3.72	
		64	390	24.83	23	7.96	
	35	38	740	23.93	23	4.06	
		55	580	23.09	23	0.38	
		72	550	25.18	23	9.48	
	50	41	870	24.07	23	4.64	
		53	755	23.21	23	0.90	
		67	680	23.54	23	2.35	

In the case of bowls, the heights of the bowls are estimated, as shown in Fig. 12. Once the height is estimated, (14) and (15) are made use of to estimate the bowl width measured at the top of the bowl (at  $Z = H$ ).

Table III presents the results for the estimation of heights of bowls. The error percentage in the height estimation of bowls

was (mean  $\pm$  std. dev) 2.75%  $\pm$  38.11%. Table IV presents the results for the estimation of diameters of bowls. The error percentage in the diameter estimation of bowls was (mean  $\pm$  std. dev) 4.58%  $\pm$  6.78%. Tables V and VI present the results from the real scenarios that were used for validation. The error percentage in the diameter/length and height estimation was

TABLE III  
ERROR IN HEIGHT ESTIMATION IN BOWLS

Object	Height (cm)	Pitch (degrees)	Dtof (mm)	Predicted (cm)	Original (cm)	Error %	Mean Error (Mean ± Std.D)
Circular Bowl (Small)	20	42	460	1.96	3	-34.60	-7.40 ± 36.45
		56	425	3.04	3	1.30	
		64	390	4.44	3	48.00	
	35	38	740	1.69	3	-43.77	
		55	580	2.13	3	-29.10	
		72	550	4.31	3	43.60	
	50	41	870	1.65	3	-45.00	
		53	755	2.26	3	-24.83	
		67	680	3.53	3	17.77	
Circular Bowl (Medium)	20	42	460	6.13	7	-12.41	-12.5 ± 40.90
		56	425	7.83	7	11.81	
		64	390	1.01	7	-85.60	
	35	38	740	5.70	7	-18.64	
		55	580	5.74	7	-17.98	
		72	550	10.77	7	53.86	
	50	41	870	3.69	7	-47.29	
		53	755	5.31	7	-24.17	
		67	680	8.95	7	27.89	
Circular Bowl (Large)	20	42	460	5.17	6	-13.87	11.67 ± 36.81
		56	425	7.95	6	32.43	
		64	390	10.32	6	72.05	
	35	38	740	4.43	6	-26.17	
		55	580	5.57	6	-7.18	
		72	550	9.85	6	64.13	
	50	41	870	4.61	6	-23.22	
		53	755	5.57	6	-7.11	
		67	680	6.84	6	13.95	

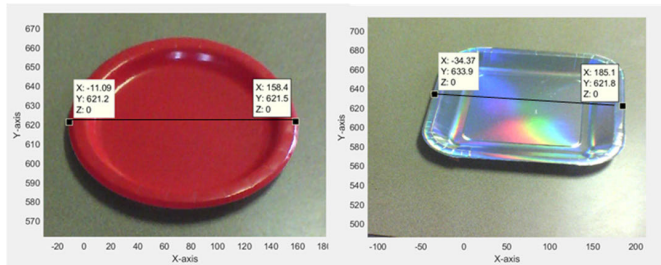


Fig. 11. Sample measurements of plate dimensions (all units are in mm).

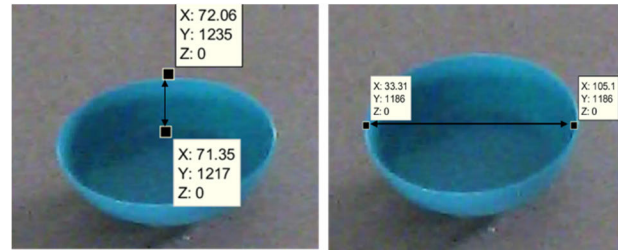


Fig. 12. Sample measurements of bowl dimensions.

(mean ± std. dev)  $-7.89\% \pm 4.71\%$  and  $4.70\% \pm 11.56\%$ , respectively.

#### IV. DISCUSSION

This study proposes a passive and automatic method for estimation of plate and bowl dimensions that involve the AIM-2 device integrated with a ToF sensor. The motivation is to use these dimensions for FPSE as in the “plate method” suggested in [17]. A geometric camera model is used to obtain real-world coordinates of the surface on which the objects of interest are present. In [1], a similar model is proposed, however, that method requires the use of a smartphone with the active participation of the user. Also, the smartphone is needed to be placed on the eating surface at a specific position.

We propose a method that does not have this requirement. We make use of a ToF ranging sensor, which can directly measure the distance between the camera and the table. The method also accounts for any lens aberrations that can cause distortions such as barrel distortion in the captured images. A major contribution of this work is the elimination of fiducial markers that have been extensively used in previous methods for FPSE. The direct measurement from the range sensor will provide the necessary dimension reference in 2-D-to-3-D model conversion.

The method makes several assumptions prior to estimation: the camera axis and the range sensor axis are parallel to each other, the roll and yaw angles of the sensor are 0, the eating surface = 0, and the walls of the bowls are flat.

TABLE IV  
ERROR IN DIAMETER ESTIMATION IN BOWLS

Object	Height (cm)	Pitch (degrees)	Dtof (mm)	Predicted (cm)	Original (cm)	Error %	Error % (Mean $\pm$ Std.D)
Circular Bowl (Small)	20	42	460	6.90	7	-1.43	4.16 $\pm$ 4.38
		56	425	6.97	7	-0.43	
		64	390	7.59	7	8.43	
	35	38	740	7.02	7	0.29	
		55	580	7.55	7	7.86	
		72	550	7.79	7	11.29	
	50	41	870	7.35	7	5.00	
		53	755	7.18	7	2.57	
		67	680	7.27	7	3.86	
Circular Bowl (Medium)	20	42	460	10.28	11	-6.55	2.52 $\pm$ 7.66
		56	425	10.26	11	-6.73	
		64	390	11.94	11	8.55	
	35	38	740	10.26	11	-6.73	
		55	580	11.64	11	5.82	
		72	550	12.66	11	15.09	
	50	41	870	11.53	11	4.82	
		53	755	11.49	11	4.45	
		67	680	11.44	11	4.00	
Circular Bowl (Large)	20	42	460	14.62	15	-2.53	7.04 $\pm$ 7.73
		56	425	14.41	15	-3.93	
		64	390	17.44	15	16.27	
	35	38	740	15.11	15	0.73	
		55	580	16.92	15	12.80	
		72	550	17.55	15	17.00	
	50	41	870	15.97	15	6.47	
		53	755	16.00	15	6.67	
		67	680	16.49	15	9.93	

TABLE V  
ERROR IN DIAMETER ESTIMATION IN CONTAINERS

Object	Predicted Diameter (cm)					Ground Truth Diameter/Length (cm)	Error (%)
	RA1	RA2	RA3	RA4	Mean		
White Box	25.70	25.18	7.17	26.37	21.11	24.50	-13.85
Pink bowl	15.10	15.75	15.10	15.39	15.33	14.20	-7.99
Green Bowl	14.83	15.20	15.41	15.25	15.17	14.90	-1.83

The proposed method was evaluated on a test bench using a calibrated protractor for positioning. Three heights and three angles were considered for testing the proposed model based on the natural behavior of participants in previous AIM-based studies. The tilt angles and the distances between the camera and the eating surface were in the selected range of pitch angles and heights.

For the measurement of bowl heights, the inner walls of the bowls were used. The rationale for using the inner walls of the bowls is that the AIM is a passive device that captures continuous images from including the start and end of meal

images, and that way the images will include an empty bowl at the end of the meal. Even if the bowl is not empty, we can measure the difference between the start and end of the meal and eventually calculate the difference in the food level. This is a major advantage of having a passive camera since there are enough images covering the entire meal. It was noticed that the error rates were higher for steep angles.

The results of estimation of dimensions for plates were acceptable with good error rates. It was noticed that the dimensions were overestimated for steep angles (70°). The estimations were most accurate for 55° compared to the other



TABLE VI  
ERROR IN HEIGHT ESTIMATION IN CONTAINERS

Object	Predicted Height (cm)					Ground Truth Height (cm)	Error (%)
	RA1	RA2	RA3	RA4	Mean		
White Box	3.46	1.563	5.66	2.98	3.42	3.8	-10.11
Pink bowl	6.185	9.62	6.7	5.841	7.09	6	18.11
Green Bowl	5.792	8.3	8.06	7.556	7.43	7	6.10

two orientations. This is a promising trend since the corresponding AIM pitch angles normally occur when a person is bending forward to grab a bite of the food in front. In addition, since the AIM captures images continuously every 10 s, there will be multiple images captured at several angles due to the forward bending of the user. The angle that typically had the lowest error rates could then be picked from the range of angles available to estimate the dimensions of the objects in the scene. This reference can then also be used for images from different orientations.

We also noticed that the error rates were lower for heights of 20 and 35 cm compared to 50 cm for the same pitch angles in the case of plate diameter estimations. This could be because the plates are more central in the images as the camera is closer, reducing the field of view (the area covered by the camera). However, for the height and diameter estimations of bowls, a height 35 cm was more accurate compared to the 20- and 50-cm cases for narrow pitch angles. The 35 cm height might be ideal for the methodology used here since the walls of the bowls are clearer to the user to mark. The best results were obtained for the heights of 20 and 35 cm at 55° pitch angles for plates and bowls, respectively.

One limitation of the study is that the bowl walls are assumed to be flat and not curved. This could be a source of error in dimension measurement and portion size estimation. The method also assumes that the plates are part of the plane  $Z = 0$ . The method does not account for the thickness of the plates or the curvature of plates. However, unlike other studies which use plates as a reference, this method is not restricted to circular plates or bowls. Any shape of plates or bowls can be included.

Finally, the proposed method was validated by wearing the AIM and collecting data for three cases and four research assistants estimated the diameters/length and heights for the same. The results suggested that except for a couple of outliers (RA3: diameter and RA2: height for white box), the estimates were reasonably accurate. Also, it should be noted that one of the cases was a hollow rectangular box. This indicates that the method could be employed for similar shaped bowls and possibly for a larger variety of bowl shapes. However, in some situations where the walls of the bowls are not flat, our assumption of the walls being flat might induce errors in estimating the height of the container accurately.

Future work could include estimating food portion sizes from the dimensions of the bowls and plates. Another possible work is to use this method to estimate the dimensions of regular-shaped foods followed by food volume. Also, the proposed method was only tested on a test bench that was stationary. Since the AIM device is primarily designed to be

mounted on the eyeglass, it is necessary to test the proposed method by mounting the sensor system on a human.

## V. CONCLUSION

In this article, we propose a wearable sensor system-based (the automatic ingestion monitor integrated with a ToF ranging sensor) method for the estimation of dimensions of plates and bowls. The contributions of this study are: 1) the model eliminates the need for fiducial markers; 2) the camera system (AIM-2) is not restricted in terms of positioning, unlike in [29] where the smartphone is required to be placed on the eating surface; 3) our model accounts for radial lens distortion in caused due to lens aberrations; 4) a distance (ToF) sensor directly gives the distance between the sensor and the eating surface; 5) the model is not restricted to circular plates; and 6) a passive method that can be used either for automatic or manual assessment of container dimensions with minimum user interaction. The error rates (mean  $\pm$  std. dev) for dimension estimation were 2.01%  $\pm$  4.10% for plate widths/diameters, 2.75%  $\pm$  38.11% for bowl heights, and 4.58%  $\pm$  6.78% for bowl diameters.

## ACKNOWLEDGMENT

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

## REFERENCES

- [1] Y. Yang, W. Jia, T. Bucher, H. Zhang, and M. Sun, "Image-based food portion size estimation using a smartphone without a fiducial marker," *Public Health Nutr.*, vol. 22, no. 7, pp. 1180–1192, May 2019, doi: [10.1017/S136898001800054X](https://doi.org/10.1017/S136898001800054X).
- [2] P. J. Stumbo, "New technology in dietary assessment: A review of digital methods in improving food record accuracy," *Proc. Nutrition Soc.*, vol. 72, no. 1, pp. 70–76, Feb. 2013, doi: [10.1017/S0029665112002911](https://doi.org/10.1017/S0029665112002911).
- [3] H. M. Al Marzooqi, S. J. Burke, M. R. Al Ghazali, E. Duffy, and M. H. S. A. Yousuf, "The development of a food atlas of portion sizes for the United Arab Emirates," *J. Food Composition Anal.*, vol. 43, pp. 140–148, Nov. 2015, doi: [10.1016/j.jfca.2015.05.008](https://doi.org/10.1016/j.jfca.2015.05.008).
- [4] H. I. Ali, C. Platat, N. El Mesmoudi, M. El Sadig, and I. Tewfik, "Evaluation of a photographic food atlas as a tool for quantifying food portion size in the United Arab Emirates," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0196389, doi: [10.1371/journal.pone.0196389](https://doi.org/10.1371/journal.pone.0196389).
- [5] R. Jayawardena and M. P. Herath, "Development of a food atlas for sri Lankan adults," *BMC Nutrition*, vol. 3, no. 1, p. 43, Dec. 2017, doi: [10.1186/s40795-017-0160-4](https://doi.org/10.1186/s40795-017-0160-4).
- [6] E. Foster, A. Hawkins, K. L. Barton, E. Stamp, J. N. S. Matthews, and A. J. Adamson, "Development of food photographs for use with children aged 18 months to 16 years: Comparison against weighed food diaries—The young person's food atlas (U.K.)," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0169084, doi: [10.1371/journal.pone.0169084](https://doi.org/10.1371/journal.pone.0169084).
- [7] M. P. Villena-Esponera, R. Moreno-Rojas, S. Mateos-Marcos, M. V. Salazar-Donoso, and G. Molina-Recio, "Validation of a photographic atlas of food portions designed as a tool to visually estimate food amounts in Ecuador," *Nutricion Hospitalaria*, vol. 36, no. 2, pp. 363–371, 2019, doi: [10.20960/nh.2147](https://doi.org/10.20960/nh.2147).

- [8] G. Turconi, M. Guarcello, F. G. Berzolari, A. Carolei, R. Bazzano, and C. Roggi, "An evaluation of a colour food photography atlas as a tool for quantifying food portion size in epidemiological dietary surveys," *Eur. J. Clin. Nutrition*, vol. 59, no. 8, pp. 923–931, Aug. 2005, doi: [10.1038/sj.ejcn.1602162](https://doi.org/10.1038/sj.ejcn.1602162).
- [9] M.-L. Ovaskainen et al., "Accuracy in the estimation of food servings against the portions in food photographs," *Eur. J. Clin. Nutrition*, vol. 62, no. 5, pp. 674–681, May 2008, doi: [10.1038/sj.ejcn.1602758](https://doi.org/10.1038/sj.ejcn.1602758).
- [10] L. Korkalo, M. Erkkola, L. Fidalgo, J. Nevalainen, and M. Mutanen, "Food photographs in portion size estimation among adolescent Mozambican girls," *Public Health Nutrition*, vol. 16, no. 9, pp. 1558–1564, Sep. 2013, doi: [10.1017/S1368980012003655](https://doi.org/10.1017/S1368980012003655).
- [11] E. Foster, J. N. Matthews, M. Nelson, J. M. Harris, J. C. Mathers, and A. J. Adamson, "Accuracy of estimates of food portion size using food photographs—The importance of using age-appropriate tools," *Public Health Nutrition*, vol. 9, no. 4, pp. 509–514, Jun. 2006, doi: [10.1079/PHN2005872](https://doi.org/10.1079/PHN2005872).
- [12] K. Nissinen et al., "Accuracy in the estimation of children's food portion sizes against a food picture book by parents and early educators," *J. Nutritional Sci.*, vol. 7, p. e35, 2018, doi: [10.1017/jns.2018.26](https://doi.org/10.1017/jns.2018.26).
- [13] V. B. Raju and E. Sazonov, "A systematic review of sensor-based methodologies for food portion size estimation," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12882–12899, Jun. 2021, doi: [10.1109/JSEN.2020.3041023](https://doi.org/10.1109/JSEN.2020.3041023).
- [14] C. Xu, Y. He, N. Khanna, C. J. Boushey, and E. J. Delp, "Model-based food volume estimation using 3D pose," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2534–2538, doi: [10.1109/ICIP.2013.6738522](https://doi.org/10.1109/ICIP.2013.6738522).
- [15] J. Dehais, M. Anthimopoulos, S. Shevchik, and S. Mougialakou, "Two-view 3D reconstruction for food volume estimation," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1090–1099, May 2017, doi: [10.1109/TMM.2016.2642792](https://doi.org/10.1109/TMM.2016.2642792).
- [16] A. Gao, F. P.-W. Lo, and B. Lo, "Food volume estimation for quantifying dietary intake with a wearable camera," in *Proc. IEEE 15th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Mar. 2018, pp. 110–113, doi: [10.1109/BSN.2018.8329671](https://doi.org/10.1109/BSN.2018.8329671).
- [17] W. Jia et al., "Imaged based estimation of food volume using circular referents in dietary assessment," *J. Food Eng.*, vol. 109, no. 1, pp. 76–86, Mar. 2012, doi: [10.1016/j.jfoodeng.2011.09.031](https://doi.org/10.1016/j.jfoodeng.2011.09.031).
- [18] M. McCrory et al., "Methodology for objective, passive, image- and sensor-based assessment of dietary intake, meal-timing, and food-related activity in Ghana and Kenya (P13-028-19)," *Current Develop. Nutrition*, vol. 3, Jun. 2019, doi: [10.1093/cdn/nzz036.P13-028-19](https://doi.org/10.1093/cdn/nzz036.P13-028-19).
- [19] S. Fang, F. Zhu, C. Jiang, S. Zhang, C. J. Boushey, and E. J. Delp, "A comparison of food portion size estimation using geometric models and depth images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 26–30, doi: [10.1109/ICIP.2016.7532312](https://doi.org/10.1109/ICIP.2016.7532312).
- [20] Z. Zhang, Y. Yang, Y. Yue, J. D. Fernstrom, W. Jia, and M. Sun, "Food volume estimation from a single image using virtual reality technology," in *Proc. IEEE 37th Annu. Northeast Bioeng. Conf.*, Apr. 2011, pp. 1–2, doi: [10.1109/NEBC.2011.5778625](https://doi.org/10.1109/NEBC.2011.5778625).
- [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 *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4545–4548, doi: [10.1109/EMBC.2012.6346978](https://doi.org/10.1109/EMBC.2012.6346978).
- [22] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, "Measuring calorie and nutrition from food image," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 8, pp. 1947–1956, Aug. 2014, doi: [10.1109/TIM.2014.2303533](https://doi.org/10.1109/TIM.2014.2303533).
- [23] J. Shang et al., "A mobile structured light system for food, volume estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Barcelona, Spain, Nov. 2011, pp. 100–101, doi: [10.1109/ICCVW.2011.6130229](https://doi.org/10.1109/ICCVW.2011.6130229).
- [24] C. J. Boushey, D. A. Kerr, J. Wright, K. D. Lutes, D. S. Ebert, and E. J. Delp, "Use of technology in children's dietary assessment," *Eur. J. Clin. Nutrition*, vol. 63, no. 1, pp. 50–57, Feb. 2009, doi: [10.1038/ejcn.2008.65](https://doi.org/10.1038/ejcn.2008.65).
- [25] N. Khanna, C. J. Boushey, D. Kerr, M. Okos, D. S. Ebert, and E. J. Delp, "An overview of the technology assisted dietary assessment project at Purdue University," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2010, pp. 290–295, doi: [10.1109/ISM.2010.50](https://doi.org/10.1109/ISM.2010.50).
- [26] F. Kong and J. Tan, "DietCam: Automatic dietary assessment with mobile camera phones," *Pervas. Mob. Comput.*, vol. 8, no. 1, pp. 147–163, Feb. 2012, doi: [10.1016/j.pmcj.2011.07.003](https://doi.org/10.1016/j.pmcj.2011.07.003).
- [27] S. Fang et al., "Single-view food portion estimation: Learning image-to-energy mappings using generative adversarial networks," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 251–255, doi: [10.1109/ICIP.2018.8451461](https://doi.org/10.1109/ICIP.2018.8451461).
- [28] C. J. Boushey, M. Spoden, F. M. Zhu, E. J. Delp, and D. A. Kerr, "New mobile methods for dietary assessment: Review of image-assisted and image-based dietary assessment methods," *Proc. Nutrition Soc.*, vol. 76, no. 3, pp. 283–294, Aug. 2017, doi: [10.1017/S0029665116002913](https://doi.org/10.1017/S0029665116002913).
- [29] M. H. Rahman et al., "Food volume estimation in a mobile phone based dietary assessment system," in *Proc. 8th Int. Conf. Signal Image Technol. Internet Based Syst.*, Nov. 2012, pp. 988–995, doi: [10.1109/SITIS.2012.146](https://doi.org/10.1109/SITIS.2012.146).
- [30] T. Bucher et al., "The international food unit: A new measurement aid that can improve portion size estimation," *Int. J. Behav. Nutrition Phys. Activity*, vol. 14, no. 1, pp. 1–11, Dec. 2017, doi: [10.1186/s12966-017-0583-y](https://doi.org/10.1186/s12966-017-0583-y).
- [31] M. Sun et al., "A wearable electronic system for objective dietary assessment," *J. Amer. Dietetic Assoc.*, vol. 110, no. 1, pp. 45–47, Jan. 2010, doi: [10.1016/j.jada.2009.10.013](https://doi.org/10.1016/j.jada.2009.10.013).
- [32] M. Sun et al., "eButton: A wearable computer for health monitoring and personal assistance," in *Proc. 51st Annu. Design Automat. Conf.*, 2014, pp. 1–6, doi: [10.1145/2593069.2596678](https://doi.org/10.1145/2593069.2596678).
- [33] A. Doulah, T. Ghosh, D. Hossain, M. H. Imtiaz, and E. Sazonov, "Automatic ingestion monitor version 2—A novel wearable device for automatic food intake detection and passive capture of food images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 568–576, Feb. 2021, doi: [10.1109/JBHI.2020.2995473](https://doi.org/10.1109/JBHI.2020.2995473).



**Viprav B. Raju** (Student Member, IEEE) received the bachelor's degree in electrical and computer engineering from Visvesvaraya Technological University (VTU), Bengaluru, India, in 2016, and the M.S. degree in electrical engineering from The University of Alabama, Tuscaloosa, AL, USA, in 2017, where he is currently pursuing the Ph.D. degree in electrical engineering.

His research interests include computer vision image processing, sensor networks, machine learning, and deep learning. His current research interests include dietary assessment and image-based food intake monitoring.



**Delwar Hossain** (Student Member, IEEE) received the bachelor's degree in electrical engineering from the Khulna University of Engineering and Technology, Khulna, Bangladesh, in 2013. He is currently pursuing the Ph.D. degree in electrical engineering with The University of Alabama, Tuscaloosa, AL, USA.

His research interests include the development of wearable systems, sensor networks, and machine learning algorithms for preventive, diagnostic, and assistive health technology, with a special focus on physical activity and dietary intake monitoring.



**Edward Sazonov** (Senior Member, IEEE) received the Diploma degree in systems engineering from the Khabarovsk State University of Technology, Khabarovsk, Russia, in 1993, and the Ph.D. degree in computer engineering from West Virginia University, Morgantown, WV, USA, in 2002.

He is currently a Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA, and the Head of the Computer Laboratory of Ambient and Wearable Systems, The University of Alabama. His research interests include wireless, ambient, and wearable devices; methods of biomedical signal processing; and pattern recognition. Devices developed in his laboratory include: a wearable sensor for objective detection and characterization of food intake; a highly accurate physical activity and gait monitor integrated into a shoe insole; a wearable sensor system for monitoring of cigarette smoking; and others. His research has been supported by the National Science Foundation, National Institutes of Health, National Academies of Science, and state agencies and private industry and foundations.