

3D Localization of Circular Feature in 2D Image and Application to Food Volume Estimation*

Wenyan Jia, Yaofeng Yue, John. D. Fernstrom, Zhengnan Zhang, Yongquan Yang, Mingui Sun

Abstract— A novel method to estimate the 3D location of a circular feature from a 2D image is presented and applied to the problem of objective dietary assessment from images taken by a wearable device. Instead of using a common reference (e.g., a checkerboard card), we use a food container (e.g., a circular plate) as a necessary reference before the volumetric measurement. In this paper, we establish a mathematical model formulating the system involving a camera and a circular object in a 3D space and, based on this model, the food volume is calculated. Our experiments showed that, for 240 pictures of a variety of regular objects and food replicas, the relative error of the image-based volume estimation was less than 10% in 224 pictures.

I. INTRODUCTION

The accurate measurement of food intake in real-world settings is extremely important in determining overall diet composition, its contribution to weight gain, and thus its role in the development of obesity and obesity-associated diseases (e.g., diabetes and heart disease). Recently, image-based dietary analysis has become an active area of research, since food photographs can be easily acquired by a hand-held device, such as a cell phone or a tablet computer. After the food items in photographs are manually or automatically identified and their volumes are determined, their caloric and nutrient contents can be calculated using on-line databases (e.g., the USDA Food and Nutrient Database for Dietary Studies [1]). Software and cell phone apps are being developed to help individuals assess their dietary habits and pursue a healthy diet [2,3]. However, such approaches usually require the active involvement of users in acquiring photos before and after meal and snack consumption, which is burdensome and may disrupt normal dietary habits. Moreover, individuals may hesitate to take pictures of foods in certain social environments. The accuracy and usefulness of the information collected may therefore be questionable. To eliminate these problems, we have developed a small, wearable device (“eButton”) to automate food intake data collection, see Fig.1 and [4]. The device is worn on a shirt or blouse and takes pictures of the scene in front of the wearer at a pre-set rate (e.g., 1 frame per second) [4,5]. All recorded images are saved to a microSD

card in the device, and later uploaded to a computer for off-line evaluation of the amounts of foods ingested. An event segmentation algorithm, together with the detection of elliptical object (e.g., the image of the dinning plate) has been developed to facilitate the manual selection of food pictures for dietary analysis [4-6].

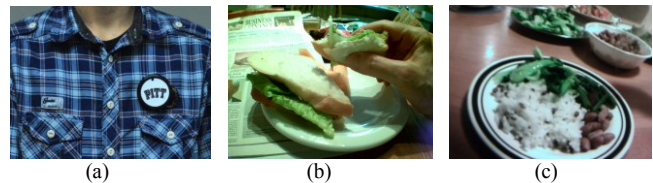


Fig. 1. (a) eButton worn naturally on chest; (b) and (c) typical pictures taken by eButton.

The calculation of the amounts of foods consumed from a 2D image requires a method for determining food volumes. It is well known that 3D volume estimation from a 2D image is impossible if no reference is provided. Typically, a checkerboard card is used for this purpose [2,3]. However, subjects may forget to place the card into the field of view at each eating episode, leading to a loss of data. A better approach would be to use a referent that is usually present in the field of view, such as a circular plate with known diameter. In the free-living environment, the diameter of the dinning plate can be measured before or after food images are taken. We therefore propose a new approach to estimate food volume by determining the 3D location of a circular plate. Although circular objects are very common in daily life, 3D localization of a circular feature from a single image is a non-trivial problem in the field of computer vision [7,8]. In 1992, a novel method was presented by Safaee-Rad and his colleagues [9], in which four transformations were used to facilitate the solution. Actually, only two transforms are necessary if the coordinate system is re-defined as in this paper. We have also proved that only one unique solution of the 3D location exists if the diameter of the circular feature is provided and the circular feature is located on a horizontal plane [10]. In our application, the plate is placed on a horizontal table when the wearer is eating a meal. So food location can be determined from one single image taken by our wearable device.

In this paper, we first introduce our algorithm to determine the 3D location of a circular object. Then, two methods to estimate food volume based on its determined location are proposed. Finally, experimental results are presented which validate the performance of our approach.

II. 3D LOCATION ESTIMATION OF A CIRCULAR OBJECT

Assuming the camera can be modeled as a pin-hole camera, the relationship between a circular feature (e.g. a

*Research supported by National Institutes of Health grant U01 HL91736.

W. Jia, Y. Yue, Z. Zhang, Y. Yang, and M. Sun are with the Department of Neurosurgery, University of Pittsburgh, PA, 15260 (e-mail: drsun@pitt.edu)

John D. Fernstrom is with Department of Psychiatry, University of Pittsburgh, PA, 15260

Y. Yang is also with Department of Computer Science, Ocean University of China, Qingdao, China.

circular dining plate) on the object plane and its image on the image plane is illustrated in Fig. 2(a). Given such a perspective projection between a circular feature in 3D space and its 2D image, a 3-D quadric cone is determined by the base (the image of the circular feature) and the vertex (the focus of the camera lens). Then the problem of determining the circular-feature's 3D orientation can be simplified as finding a plane that intersects the cone in a circle.

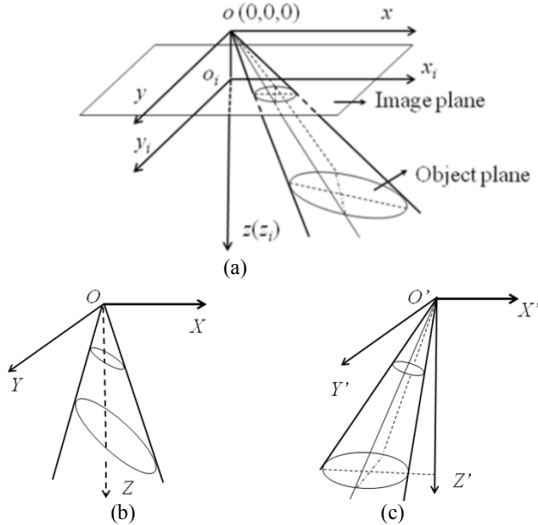


Fig. 2. Geometric model of the camera perspective system and schematic diagram of rotational transformations. Origin o is the focus of the camera lens. $f=|oo_i|$ is the focal length of the camera. xyz is the camera coordinate system and $x_iy_iz_i$ is the image coordinate system. oz is the optical axis of the camera. ox is parallel to the horizontal axis of the image. (a) is the perspective relationship between the circular feature and its image with respect to the camera coordinate system. (b) is the coordinate system after transformation \mathbf{T}_1 in which the cone is centralized. (c) is another coordinate system after transformation \mathbf{T}_2 in which the Z' is perpendicular to the object plane.

In the standard coordinate system of camera xoy defined in Fig. 2(a), the base of the cone is defined by the image of the circular feature in the image plane (i.e., an ellipse). Its equation can be described as:

$$a'x^2 + 2h'xy + b'y^2 + 2g'x + 2f'y + d' = 0 \quad (1)$$

$$z = f$$

where f is the focal length of the camera.

In this coordinate system, the equation of the cone can be formulated in the following quadratic surface function [11]:

$$a'f^2x^2 + 2h'f^2xy + b'f^2y^2 + 2g'fxz + 2f'fyz + d'z^2 = 0 \quad (2)$$

which can also be written as:

$$[x \ y \ z] \cdot \mathbf{Q} \cdot [x \ y \ z]^T = 0 \quad (3)$$

$$\text{where } \mathbf{Q} = \begin{bmatrix} a'f^2 & h'f^2 & g'f \\ h'f^2 & b'f^2 & f'f \\ g'f & f'f & d' \end{bmatrix}$$

The object plane where the circular feature locates can be defined as [11]:

$$lx + my + nz = p \quad (4)$$

i.e., $[l \ m \ n] \cdot [x \ y \ z]^T = p$, where (l, m, n) represents the orientation of the object plane with $l^2 + m^2 + n^2 = 1$ and p is a constant. Thus, determining the orientation

of the circular feature is to solve (l, m, n) under the condition that the intersection between the quadric cone and the object plane is a circle.

In order to solve this problem in a simpler way, two rotational transformations are applied. The transform \mathbf{T}_1 is utilized to centralize the quadric cone, see Fig.2 (b). The relationship between the xyz and XYZ coordinate becomes:

$$[x \ y \ z]^T = \mathbf{T}_1 \cdot [X \ Y \ Z]^T. \quad (5)$$

If the three eigen values of matrix \mathbf{Q} are defined as $\lambda_1, \lambda_2, \lambda_3$, \mathbf{T}_1 can be calculated as:

$$\mathbf{T}_1 = \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix}. \quad (6)$$

where $[l_i \ m_i \ n_i]^T$ is the eigen vector corresponding to the eigen value $\lambda_i, i=1,2,3$ [11]. The estimated values for l_i, m_i, n_i must satisfy the right-hand rule.

After applying this transformation, the cone equation (3) becomes:

$$[X \ Y \ Z] \cdot \mathbf{T}_1^T \cdot \mathbf{Q} \cdot \mathbf{T}_1 \cdot [X \ Y \ Z]^T = 0 \quad (7)$$

$$\text{i.e., } [X \ Y \ Z] \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot [X \ Y \ Z]^T = 0 \quad (8)$$

In the new coordinate system XYZ , the equation of the object plane (4) can be re-written as:

$$[L \ M \ N] \cdot [X \ Y \ Z]^T = p \quad (9)$$

where (L, M, N) represents the orientation of the object plane and

$$L^2 + M^2 + N^2 = 1. \quad (10)$$

So the relationship between the (l, m, n) and (L, M, N) can be easily derived as:

$$[l \ m \ n] = [L \ M \ N] \cdot \mathbf{T}_1^{-1}. \quad (11)$$

In this system, the cone equation has been simplified, but the plane equation has not. Another transformation \mathbf{T}_2 is performed to simplify the equation of the object plane by making the Z' axis normal to the object plane, see Fig.2(c).

$$[X \ Y \ Z]^T = \mathbf{T}_2 \cdot [X' \ Y' \ Z']^T \quad (12)$$

$$\text{where } \mathbf{T}_2 = \begin{bmatrix} -\frac{M}{\sqrt{L^2+M^2}} & -\frac{LN}{\sqrt{L^2+M^2}} & L \\ \frac{L}{\sqrt{L^2+M^2}} & -\frac{MN}{\sqrt{L^2+M^2}} & M \\ 0 & \sqrt{L^2+M^2} & N \end{bmatrix}. \quad (13)$$

In the new coordinate system $X'Y'Z'$, the equation of the object plane becomes:

$$Z' = p \quad (14)$$

Substituting (12) and (14) into (8), we can get the intersection curve between the quadratic cone and the object plane as:

$$[X' \ Y' \ p] \cdot \mathbf{T}_2^T \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot \mathbf{T}_2 \cdot \begin{bmatrix} X' \\ Y' \\ p \end{bmatrix} = 0 \quad (15)$$

$$\text{i.e., } AX'^2 + BX'Y' + CY'^2 + DX' + EY' + F = 0 \quad (16)$$

$$\begin{aligned} \text{where } A &= \frac{\lambda_1 M^2}{L^2 + M^2} + \frac{\lambda_2 L^2}{L^2 + M^2} \\ B &= \frac{2LMN}{L^2 + M^2} (\lambda_1 - \lambda_2) \\ C &= \frac{\lambda_1 L^2 N^2}{L^2 + M^2} + \frac{\lambda_2 M^2 N^2}{L^2 + M^2} + \lambda_3 (L^2 + M^2) \\ D &= \frac{2pLM}{\sqrt{L^2 + M^2}} \\ E &= \frac{2pN}{\sqrt{L^2 + M^2}} (-\lambda_1 L^2 - \lambda_2 M^2 + \lambda_3 (L^2 + M^2)) \\ F &= p^2 (\lambda_1 L^2 + \lambda_2 M^2 + \lambda_3 N^2) \end{aligned}$$

The necessary and sufficient condition that (16) represents a circle is: $A = C$, $B = 0$

$$\text{i.e.: } \frac{\lambda_1 M^2}{L^2 + M^2} + \frac{\lambda_2 L^2}{L^2 + M^2} = \frac{\lambda_1 L^2 N^2}{L^2 + M^2} + \frac{\lambda_2 M^2 N^2}{L^2 + M^2} + \lambda_3 (L^2 + M^2) \quad (17)$$

$$\frac{2LMN}{L^2 + M^2} (\lambda_1 - \lambda_2) = 0 \quad (18)$$

Combining (17) and (18) with (10), three equations with three unknowns are obtained. Although there are four symmetrical solutions with respect to the origin of $X'Y'Z'$ frame, it can be uniquely determined in our real application since the camera was worn by a participant and capture pictures when he/she is eating beside a table. In this case, it has been shown that the solutions of L , M , N are in the following form according to the defined coordinate in Fig. 2(a) [10]:

$$L = 0, \quad M = \frac{\sqrt{\lambda_2 - \lambda_1}}{\sqrt{\lambda_2 - \lambda_3}}, \quad N = \frac{\sqrt{\lambda_1 - \lambda_3}}{\sqrt{\lambda_2 - \lambda_3}}, \quad L^2 + M^2 \neq 0. \quad (19)$$

Then the plate orientation (l, m, n) can be calculated according to (11). If $L=M=0$, it means that the object plane is perpendicular to the optical axis of the camera and the center of the circle is located at the optical axis. In this case, the image of the circular feature is still a circle, and its location can be derived easily, if the diameter of the circular feature is known.

In the $X'Y'Z'$ system, Z' axis is perpendicular to the object plane. Then, the z -distance of the object plane has been derived as [9]:

$$p = \frac{c_1 r}{\sqrt{c_2^2 + c_3^2 - c_1 c_4}} \quad (20)$$

Here r is the radius of the circular feature and $C_1 - C_4$ are parameters determined by the eigen vectors and eigen values of \mathbf{Q} :

$$\begin{aligned} C_1 &= \lambda_1 l_1^2 + \lambda_2 l_2^2 + \lambda_3 l_3^2 \\ C_2 &= \lambda_1 l_1 n_1 + \lambda_2 l_2 n_2 + \lambda_3 l_3 n_3 \\ C_3 &= \lambda_1 m_1 n_1 + \lambda_2 m_2 n_2 + \lambda_3 m_3 n_3 \\ C_4 &= \lambda_1 n_1^2 + \lambda_2 n_2^2 + \lambda_3 n_3^2 \end{aligned} \quad (21)$$

III. FOOD VOLUME ESTIMATION FROM A SINGLE IMAGE

To estimate food volume, we must define several simple geometric shapes to model food shape (e.g., cuboid, sphere). The volumes of such geometric shapes can be determined by dimensional measurements (e.g., length, width, height, and/or diameter). We propose two methods to calculate the volume of these geometric shapes (see Fig. 3). The first

method, called the ‘‘point-clicking’’ method, involves manually clicking points in the image (the red points in Fig. 3(a)) to represent the length, width and height of the object. The length of each dimension is then calculated using derived equations to estimate the distance of two arbitrary points parallel/perpendicular to the plane of the plate [10]. The second method is called the ‘‘wireframe-fitting’’ method. A 3D virtual wireframe in the camera coordinate system is projected onto the 2D plate plane, and the projection in the image is displayed (see Fig. 3(b)). The operator then manually adjusts (drags or zooms) the projection to fit the shape of the food item. Since the volume of this fitted wireframe is known, the volume of the food item can be estimated [12].



Fig. 3. The measurement of geometric dimensions in images. (a) ‘‘point-clicking’’ method; (b) ‘‘wireframe-fitting’’ method.

IV. EXPERIMENTS AND RESULTS

Two experiments were conducted to validate the feasibility and accuracy of the volume estimation. One regular-shaped object (i.e., LEGO block) and seven food models (Nasco, Fort Atkinson, Wisconsin) were selected as samples to represent different shapes. We used a cuboid to model a LEGO block, a piece of bread and a block of cornbread, a cylinder to model a hamburger, a cone for a glass of milk and half grapefruit, a sphere for an orange, and a half-sphere for a serving of ice cream. The diameter of the circular plate used in these experiments was 260mm. Since the volumes of the food replicas provided by the manufacturer were not accurate, the true volume for each food model (except milk) was determined as the average value over three measurements using the water displacement method [13]. A webcam (Logitech Pro 9000) was used to take pictures of all the samples at 2M-pixel resolution. This camera’s intrinsic parameters were pre-calibrated using a checkerboard pattern. Typical pictures are shown in Fig.4. For each picture, a plate detection algorithm was first used to automatically find the boundary of circular plate [6]. Detection results are also shown in Fig. 4. Then an operator was asked to manually select a shape model for the food item and then measure the food volume using our method.



Fig. 4. Typical pictures taken in the experiment and the detected boundary of the plate shown as a red ellipse in each picture.

A. Volume estimation for a regularly-shaped object

In the first experiment, 107 pictures were taken at 107 different locations for a LEGO block and measured by the “point-clicking” method. For this block, clicking four points is enough for calculating its volume, which is much simpler than manually dragging and fitting it with a virtual wireframe. To identify the error introduced by manually selecting points, two operators were asked to estimate the volume of the LEGO. The mean distance between the camera’s optical center and the plate center over the 107 locations was 40.8cm (range: 25.1~68.4cm). Table I shows the results for one expert and one novice. The mean relative error was less than 2%. A student t-test showed there to be no significant difference between measurements made by the two operators.

Table I: Relative error estimated by an expert and a novice

	Ground Truth	Relative Error (Expert)	Relative Error (Novice)
Bottom Area	10.14cm ²	-1.96% ± 3.69%	-2.79% ± 4.16%
Height	7.68cm	0.46% ± 1.41%	0.84% ± 1.81%
Volume	77.95cm ³	-1.52% ± 3.64%	-1.96% ± 4.70%

B. Volume estimation for food replicas

In the second experiment, thirty pictures were taken for each food replica, as well as the LEGO block, at thirty different locations. Their volumes were estimated by both the “point-clicking” and “wireframe-fitting” methods. A total of 240 pictures were tested in our study. The estimated errors for all the samples are listed in Table II and Fig. 5. For the LEGO block, similar accuracy was obtained for each method. However, for the food replicas, the “wireframe-fitting” method produced much better results. For all 240 pictures, the errors in the estimated volumes were less than 25% in 209 pictures using the “point-clicking method”, and less than 10% in 224 pictures using the “wireframe-fitting” method. From Fig. 5, we can also see that the “wireframe-fitting” method is much more robust than the “point-clicking” method, especially for the food items whose dimensional length is difficult to find (e.g., hamburger). The error for the volume of bread is relative large because of its low height (about 1.2cm) and irregular border shape. Notably, the height estimation is difficult to make accurately when the angle between the camera optical axis and the table is large (see bread image in Fig. 4).

Table II: Relative error for food replicas and a LEGO block

	Ground Truth	Point-Clicking Method	Wireframe-Fitting Method
LEGO	77.95 cm ³	-2.92% ± 3.10%	-1.32% ± 2.22%
Cornbread	93 cm ³	-20.43% ± 9.98%	0.99% ± 2.40%
Orange	152 cm ³	-15.63% ± 6.20%	0.10% ± 2.86%
Hamburger	308 cm ³	14.75% ± 19.35%	0.31% ± 6.32%
Bread	106 cm ³	15.01% ± 28.50%	2.85% ± 8.56%
Milk	240 cm ³	-1.45% ± 8.33%	2.49% ± 1.84%
Grapefruit	272 cm ³	-0.27% ± 12.35%	-4.31% ± 4.10%
Ice Cream	80 cm ³	-19.22% ± 4.6%	-6.58% ± 4.28%

V. CONCLUSIONS

In this paper, we have introduced a novel method to calculate the 3D location of a circular feature and applied this method to the estimation of food volume. We have also

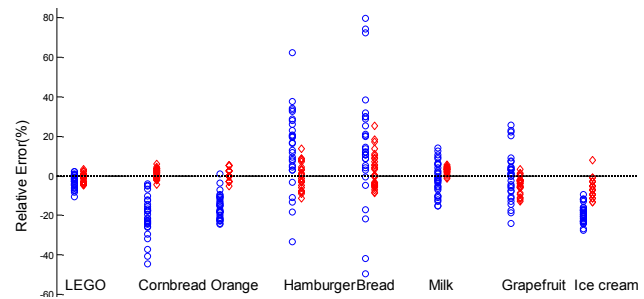


Fig. 5. The distribution of relative errors for all the testing samples. Each sample was tested at thirty different locations. The blue circle represents the relative error estimated by point-clicking method, and the red diamond represents the error estimated by wireframe-fitting method.

developed two model-based approaches to estimate food volume from a single input image given the 3D food locations. A cuboid and seven food replicas on a circular plate have been used in our experiments for performance evaluation. Our results have indicated that, using the “wireframe-fitting” method, the average volume estimation error was less than 10%, calculated using 224 input pictures. This represents a significant improvement over the current self-reporting method which is inaccurate and unreliable.

REFERENCES

- [1] USDA Food and Nutrient Database for Dietary Studies, 2010. Beltsville, MD: Agricultural Research Service, Food Surveys Research Group.
- [2] F. E. Thompson, and A. F. Subar, “Chapter 1. dietary assessment methodology,” in *Nutrition in the Prevention and Treatment of Disease*, A. M. Coulston, C. L. Rock, and E. R. Mosen, Eds. San Diego, CA: Academic Press, 2001, pp.3-30
- [3] F. Zhu, M. Bosch, I. Woo, S. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp, “The use of mobile devices in aiding dietary assessment and evaluation,” *IEEE J Sel Top Signal Process*, vol. 4, pp. 756-766, 2010.
- [4] eButton: a wearable computer developed by the University of Pittsburgh, available at www.lcn.pitt.edu/eButton
- [5] M. Sun, J. D. Fernstrom, W. Jia, S. A. Hackworth, N. Yao, Y. Li, C. Li, M. H. Fernstrom, and R. J. Sclabassi, “A wearable electronic system for objective dietary assessment,” *J Am Diet Assoc*, vol. 110, pp. 45-7, 2010.
- [6] J. Nie, Z. Wei, W. Jia, L. Li, J. D. Fernstrom, R. J. Sclabassi, and M. Sun, “Automatic detection of dining plates for image-based dietary evaluation,” in *Conf Proc IEEE Eng Med Biol Soc*, Buenos Aires, Argentina, 2010, pp. 4312-4315.
- [7] Y. C. Shiu, S. Ahmad, “3D location of circular and spherical features by monocular model-based vision,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Cambridge, MA, 1989, vol.2, pp. 576 – 581.
- [8] R. D. Sheu, A. H. Bond, “A generalized method for 3D object location from single 2D images,” *Pattern Recogn.*, vol. 25, no. 8, pp. 771-786, 1992.
- [9] R. Safaee-Rad, I. Tchoukanov, K. C. Smith, and B. Benhabib, “Three-dimensional location estimation of circular features for machine vision,” *IEEE Trans.Robot. Autom.*, vol. 8, no. 5, pp. 624-640, 1992.
- [10] W. Jia, Y. Yue, J.D. Fernstrom, N. Yao, R.J. Sclabassi, M.H. Fernstrom, M. Sun, “Image-based estimation of food volume using circular referents in dietary assessment,” *J. Food Eng.*, vol. 109, no. 1, 2012, pp. 76-86.
- [11] R. J. T. Bell, *An elementary treatise on coordinated geometry of three dimensions* (3rd ed.). London: Macmillan, UK, 1944.
- [12] Z. Zhang, Y. Yang, Y. Yue, J.D. Fernstrom, W. Jia W, M. Sun, “Food volume estimation from a single image using virtual reality technology”. in *Proc. IEEE 37th Annual Northeast Biomedical Engineering Conference*; 2011 April 1-3; Troy, NY.
- [13] S. W. Hughes. “Archimedes revisited: a faster, better, cheaper method of accurately measuring the volume of small objects”. *Physics Education*, vol. 40, no.5, 2005, pp. 468-474.