

Published in final edited form as:

*Proc IEEE Annu Northeast Bioeng Conf.* 2012 December 31; 2012: 95–96. doi:10.1109/NEBC.2012.6206979.

## 3D/2D Model-to-Image Registration for Quantitative Dietary Assessment

Hsin-Chen Chen<sup>1,2</sup>, Wenyan Jia<sup>2</sup>, Zhaoxin Li<sup>2,4</sup>, Yung-Nien Sun<sup>1</sup>, and Mingui Sun<sup>2,3</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan, R.O.C

<sup>2</sup>Department of Neurosurgery, University of Pittsburgh, Pittsburgh, PA 15213, USA

<sup>3</sup>Department of Electrical & Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15213, USA

<sup>4</sup>School of Computer Science and Technology, Harbin Institute of Technology, China

### Abstract

Image-based dietary assessment is important for health monitoring and management because it can provide quantitative and objective information, such as food volume, nutrition type, and calorie intake. In this paper, a new framework, 3D/2D model-to-image registration, is presented for estimating food volume from a single-view 2D image containing a reference object (i.e., a circular dining plate). First, the food is segmented from the background image based on Otsu's thresholding and morphological operations. Next, the food volume is obtained from a user-selected, 3D shape model. The position, orientation and scale of the model are optimized by a model-to-image registration process. Then, the circular plate in the image is fitted and its spatial information is used as constraints for solving the registration problem. Our method takes the global contour information of the shape model into account to obtain a reliable food volume estimate. Experimental results using regularly shaped test objects and realistically shaped food models with known volumes both demonstrate the effectiveness of our method.

### I. Introduction

Obesity has become a major health problem in many parts of the world because it is usually associated with an increased risk of cardiovascular disease, diabetes, and other chronic diseases. To study overweight and obesity, monitoring individual's food intake is important. Traditionally, dietary evaluation is conducted by self-report, which is subjective and inaccurate. Recently, image-based food volume estimation has been studied by a number of investigators. Chae et al. [1] and Zhu et al. [2] presented model-based methods, which achieved good results by an average error less than 15%. However, their methods only utilized 2D feature points to estimate the 3D model volume. In cases where a food has a complex shape, it is difficult to detect a sufficient number of feature points accurately from a single-view image in order to infer the volume using a 3D model. As a result, their methods are sensitive to feature detection errors, and are suitable only in limited cases.

In this paper, we propose a new framework based on a model-to-image registration process for estimating food volume from a single-view image. Instead of relying on feature points,

we use global food shape properties to obtain a more reliable estimate. Moreover, the spatial information of the circular plate obtained from the camera calibration process was adopted to reduce the solution range of the registration cost function. Our initial experiments indicate that this method is more robust in food volume estimation.

## II. 3D/2D Model-to-Image Registration

Given a single-view 2D food image (see Fig. 1(a)), the first step is to segment the food item from the background region. Next, a 3D model is manually chosen to represent the food shape and is projected onto the image. Then, the position, orientation and scale of the model are optimized automatically by model-to-image registration. Finally, the food volume is estimated from the volume of the registered 3D model.

In the segmentation step, Otsu's thresholding [3] is applied to the saturation channel resulting from the hue-saturation-intensity transformation (HIS). The procedure reduces the shadow effect in the image. In Otsu's thresholding, the optimal threshold  $k^*$ , which separates two classes, is obtained by using optimization:

$$\arg \max_{k^*} \{ \sigma_B^2(k), 0 \leq k \leq 255 \} \quad (1)$$

where  $\sigma_B^2 = \omega_0 \omega_1 (\mu_0 - \mu_1)^2$  is the inter-class variance, and  $w_0$ ,  $w_1$ ,  $\mu_0$ , and  $\mu_1$  are the probabilities of class occurrences and the mean levels of the two classes, respectively. An example of segmentation result using Otsu's thresholding is shown in Fig. 1(b). As the result may contain small holes, morphological closing is performed. A convex hull tightly enclosing the food regions is built by Andrew's monotone chain algorithm [4]. The final segmentation result is shown in Fig. 1(c).

In order to estimate the food volume, we propose a 3D/2D model-to-image registration scheme to fit a 3D virtual model onto the food item in the image. Fig. 2 illustrates this concept. First, a 3D model is manually selected and is generated in the camera projection space. Let  $\mathbf{S}$  be the set of surface points of the model with  $\Theta = (t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, s_x, s_y, s_z)$  as its pose parameters, including three displacements, three rotation angles and three scales with respect to the camera coordinates. We can project the model onto the image plane by using a perspective projection matrix  $\mathbf{M}$  derived by camera calibration [5]. The projection points are obtained by  $\mathbf{S}' = \mathbf{M} \cdot \mathbf{T}(\Theta) \cdot \mathbf{S}$ , where  $\mathbf{T}$  is a  $4 \times 4$  transformation matrix which depends on the pose parameters. The 3D/2D registration is to find the optimal pose parameters  $\Theta$  of the 3D model such that the projected contour best matches the food contour by minimizing a cost function  $F_c$ :

$$F_c = \frac{1}{m} \sum_j^m I_{DT}(\mathbf{v}_j) \quad (2)$$

where  $\mathbf{v}_j \in \mathbf{S}'$  represents the  $j$ -th projected model contour point,  $m$  is the number of projected contour points, and  $I_{DT}$  is the distance (in terms of number of pixels) between the closest points at the food contour and the projected model contour.

When (2) is implemented, a problem may arise since a 3D model with different pose parameters may produce similar projection contours in a single-view image, yielding ambiguous solutions. Let us make a reasonable assumption that there is a reference object, such as a dining plate, in the image. Firstly, we derive the food location in the camera coordinate system using the circular plate (shown in Fig. 1(a)) [1]. The three displacements,  $t_x$ ,  $t_y$ , and  $t_z$ , can then be represented using only two displacements  $t_{x1}$  and  $t_{y1}$  on the x-y plane of the camera coordinates:

$$\begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T = R(\theta_P, \mathbf{n}_P) \begin{bmatrix} t_{x1} & t_{y1} & 0 \end{bmatrix}^T \quad (3)$$

where  $R$  is the rotation matrix,  $\theta_P$  is the angle between the norm of the plate and the z-axis of the camera coordinates, and  $\mathbf{n}_P$  is the outer product of them. In our case, the three angles  $\theta_x$ ,  $\theta_y$ , and  $\theta_z$  can be replaced with one rotation angle around the plate norm. In addition,  $s_x$ ,  $s_y$ , and  $s_z$  can be simplified depending on whether the selected model needs isotropic or non-isotropic scale adjustment. Consequently, the dimension of pose parameters of a 3D model can be reduced (less than or equal to six) to accelerate the registration process. Fig. 3 shows the projection contours of models before and after registration, respectively.

### III. Results and Discussion

In our experiments, a Lego brick and four food replicas including an orange, an ice cream, a grapefruit and a hamburger were used to validate the proposed method. Their true sizes were obtained by either manual measurement or water displacement. The relative error  $E$  was defined as

$$(V_E - V_G) \cdot 100\% / V_G \quad (4)$$

where  $V_E$  is the estimated value and  $V_G$  is the ground truth. Five images from each item were acquired with a Logitech webcam (Pro 9000, Logitech) at different locations. At each location, the food volume is estimated. The relative errors of the estimated volumes are listed in Table I. The estimation error for Lego is the smallest because of its simple shape. For the objects with non-regular shape, our method achieved a satisfactory accuracy with an average error of 10% based on single image inputs, much better than the volume estimation by human eyes, which is commonly utilized in current dietary studies.

### IV. Conclusion

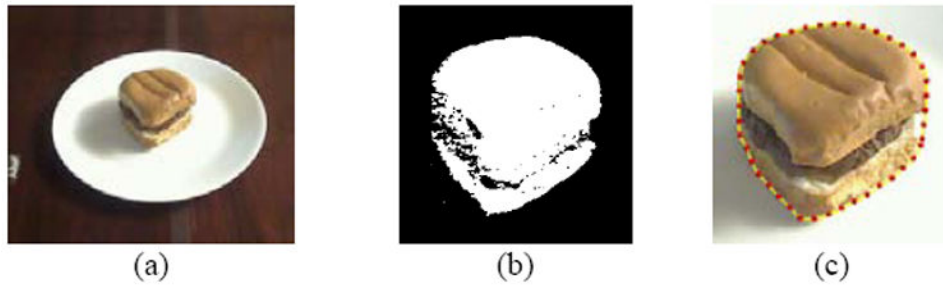
We have proposed a new framework to register a 3D model to the food item in a single-view 2D image, and applied this method to food volume estimation. Our preliminary experiments have demonstrated that this framework can provide a satisfactory accuracy despite the incomplete volumetric information in the 2D image. More importantly, this method can be generalized to common foods as long as a food can be fitted reasonably well by a regular shape (e.g., an ellipsoid or a wedge). In the future, we plan to combine different models to further improve estimation results.

## Acknowledgments

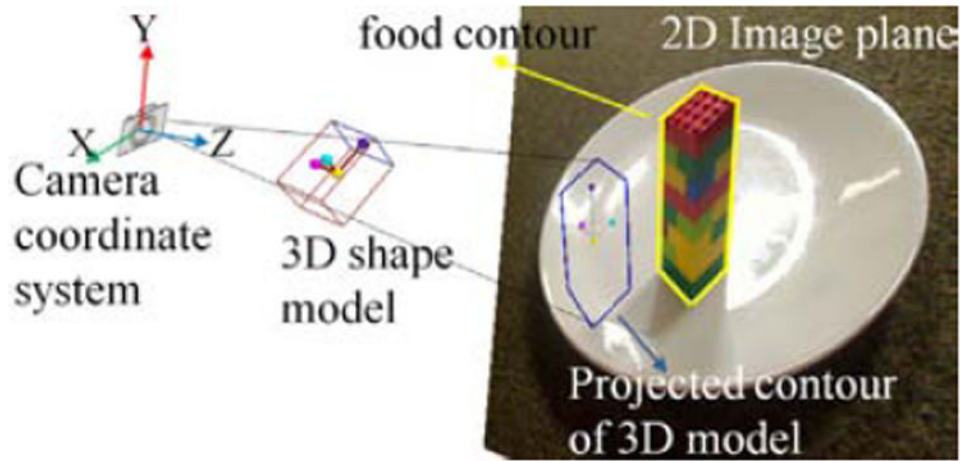
This work was supported in part by the National Institutes of Health grant U01 HL91736 and a scholarship support from the National Science Council, Taiwan, R.O.C. awarded to the first author.

## References

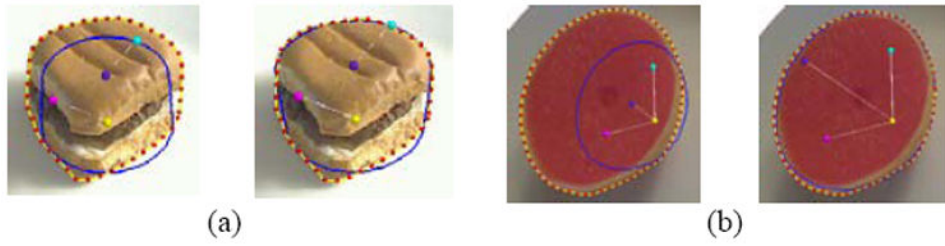
1. Chae J, Woo I, Kim SY, Maciejewski R, Zhu F, Delp EJ, Boushey CJ, Ebert DS. Volume estimation using food specific shape templates in mobile image-based dietary assessment. *Proc SPIE*. 2011; 7873:78730K.
2. Zhu F, Bosch M, Woo I, Kim SY, Boushey CJ, Ebert DS, Delp EJ. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE J Sel Top Signa*. 2010; 4:756–766.
3. Otsu N. A threshold selection method from gray-Level histograms. *IEEE Trans Sys Man Cyber*. 1979; 9:62–66.
4. Andrew AM. Another efficient algorithm for convex hulls in two dimensions. *Inform Process Lett*. 1979; 9:216–219.
5. Zhang Z. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell*. 2000; 22:1330–1334.
6. Jia W, Yue Y, Fernstrom JD, Yao N, Sclabassi RJ, Fernstrom MH, Sun M. Image based estimation of food volume using circular referents in dietary assessment. *J Food Eng*. 2012; 109:76–86. [PubMed: 22523440]



**Fig. 1.** Food segmentation: (a) original image; (b) segmentation result using Otsu's thresholding; (c) final segmentation result.



**Fig. 2.**  
Illustration for the 3D/2D model-to-image registration.



**Fig. 3.** Two examples of projected contours of 3D models before and after registration, respectively. Blue contours are the projection from 3D models, and the red ones are from food segmentation.

**Table 1**

Relative Error of Food volume estimation (%).

Sample	Lego		Orange		Ice cream		Grapefruit		Hamburger	
	Cuboid	Sphere	Half-sphere	Half-sphere	Half-sphere	Half-sphere	Half-sphere	Cylinder		
Image 1	-1.89	-1.01	5.74	9.43	-4.66					
Image 2	0.45	1.23	18.64	-9.39	-1.36					
Image 3	0.94	8.02	1.61	13.88	-9.00					
Image 4	4.47	-0.30	3.55	2.36	6.77					
Image 5	2.39	8.96	-3.31	3.81	5.67					
Mean	1.27	3.38	5.25	4.02	-0.52					
Std	2.36	4.75	8.20	8.79	6.73					