Volume Estimation Using Food Specific Shape Templates in Mobile Image-Based Dietary Assessment

Junghoon Chae[a], **Insoo Woo**[a], **SungYe Kim**[a], **Ross Maciejewski**[a], **Fengging Zhu**[a], **Edward J. Delp**[a], **Carol J. Boushey**[b], and **David S. Ebert**[a]

[a]School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana USA

[b]Department of Foods and Nutrient, Purdue University, West Lafayette, Indiana USA

Abstract

As obesity concerns mount, dietary assessment methods for prevention and intervention are being developed. These methods include recording, cataloging and analyzing daily dietary records to monitor energy and nutrient intakes. Given the ubiquity of mobile devices with built-in cameras, one possible means of improving dietary assessment is through photographing foods and inputting these images into a system that can determine the nutrient content of foods in the images. One of the critical issues in such the image-based dietary assessment tool is the accurate and consistent estimation of food portion sizes. The objective of our study is to automatically estimate food volumes through the use of food specific shape templates. In our system, users capture food images using a mobile phone camera. Based on information (i.e., food name and code) determined through food segmentation and classification of the food images, our system choose a particular food template shape corresponding to each segmented food. Finally, our system reconstructs the three-dimensional properties of the food shape from a single image by extracting feature points in order to size the food shape template. By employing this template-based approach, our system automatically estimates food portion size, providing a consistent method for estimation food volume.

Keywords

Food volume estimation; shape templates; dietary assessment; mobile image

1. INTRODUCTION

With growing concerns about adolescent and adult obesity and other health problems related to diet,[1, 2] education programs for obesity prevention have been developed to inform the general populace of health risks of being overweight and encourage healthy eating patterns.[3] One such means of tracking and analyzing eating patterns is through dietary assessments. Dietary assessment methods provide researchers with valuable information needed for assessment and hypothesis generation regarding dietary imbalance.[4] The dietary assessment methods include recording, cataloging and analyzing daily dietary records to monitor energy and nutrient intakes in a self-monitoring environment. One means of improving the data collected in these self assessed environments is through the use of technology. Technology assisted assessment is valid, feasible and promising; it can improve dietary assessment quality, and reduce the burden on both users and dietitians.[5, 6]

Given the ubiquity of mobile devices with built-in cameras, one possible means of improving dietary assessment is through photographing foods and inputting these images into a system that can determine the nutrient content of foods in the images.[7, 8] Such image-based nutrient analysis methods have focused on food identification and classification,[9] whereas little research has been done on automated food volume estimation because of the difficulty of obtaining accurate volume estimates from a single image.[10] In addition, image-based food volume estimation is highly affected by the prerequisite image analysis, such as the segmentation of food regions and extraction of geometric parameters to reconstruct 3D shapes. Particularly, false segmentation caused by shadows and reflections in an image and noise along the segmentation boundaries can seriously deteriorate the accuracy of volume estimates.

In our previous work,[11] we proposed an automated volume estimation method for approximating food volumes with 3D primitive shapes reconstructed from a single image and demonstrated that volume estimates can be improved through user interaction. Compared to our previous work, this study employs food specific shape templates to support a variety of food shapes. Furthermore, we improve the accuracy of volume estimation by minimizing the false-segmented regions and smoothing the segmentation boundaries of foods.

2. RELATED WORK

Previously, many approaches have tried to solve shape matching problems: measuring the similarity and finding a set of correspondences between shapes using geometric information.[12–15] In template based matching approaches, Berg and Malik[14] try to solve the problem of finding points correspondences in images through robust template matching. Gavrila[15] employed a template based tree structure to take into account coarse and fine level shape matching. For local feature based shape matching methods, Belongie et al.[12] considers neighborhood points on the contour of a shape for efficient 2D shape matching. However, the purely geometric information used in the previous template based matching works is not sufficient for volume estimation of complex real world food-shapes. To recover deficient information, our shape template approach allows us to apply multiple feasible algorithms and various adjustable methods.

Geometric shape analysis methods can be classified according to different criteria. Pavlidis[16] propose two classifications: the use of the interior (global) and the boundary (local) of a shape as opposed to the interior of the shape. The interior and boundary shape analysis methods include medial axis and mathematical morphology, respectively. In this work both the medial axis and mathematical morphology methods are used to increase the accuracy of volume estimation. Mathematical morphology, as a set of mathematical tools for image analysis, has been used for geometric manipulation of boundaries of an object.[17, 18] Morphological opening and closing are useful in the smoothing of not only binary images (the methods were originally developed based on binary images), but also multi-type images.[19–21] Peters[22] describes a new morphological image cleaning algorithm for noise reduction, while preserving thin features in gray-scale images. Yang et al.[23] propose a method for image processing including dilation, erosion and edge extraction and noise reduction in binary morphology. We use binary mathematical morphology to make a segmented image appear smooth for removing false information in an image.

In computer vision, some approaches have proposed hierarchical skeletal shape descriptions for topological shape matching using the medial axis.[24–26] Ho et al.[27] utilized the medial axis for shape smoothing. In our implementation the medial axis is used for global geometric feature extraction to support robust volume computation.

3. FOOD SPECIFIC SHAPE TEMPLATE FOR FOOD VOLUME ESTIMATION

The goal of our work is to obtain the most accurate volume estimation of food from a single image using shape templates while minimizing user interactions. The underlying concept of our template based approach is to factor in the geometric properties of the segmented region using specific template shapes associated with the extracted food name/code.

Our shape template based approach consists of camera calibration, false information removal, feature extraction, and 3D shape reconstruction to estimate food volumes as illustrated in Figure 1. We use two images (one is a meal image taken by the user and the other is the segmented region of a food obtained using Zhu et al.'s[10] approach) and food information (food name and code obtained from image segmentation and classification methods[10]) as inputs. In the camera calibration step, using the original food image including a fiducial marker, camera parameters are calibrated to reconstruct an actualsized 3D shape of a food.[11] Given the food name and code from the food classification, we associate each food with a template shape. For example, a glass of milk would correspond to a generalized cylindrical shape, an orange to spherical one, etc. Once we find a best-matched template shape, we minimize the false information in the segmented region to improve the next step, feature extraction. We extract feature points from the optimized segmented region using our feature point detection algorithm and shape analysis techniques: medial axis and active contour. The extracted feature points are then used to determine geometric information for a food, such as the height, radius and area. Finally, we reconstruct the 3D shape to compute the food volume using geometric information determined in the previous step.

By applying this template-based approach, we are able to automatically estimate portion sizes from food images. This reduces the burden of users having to estimate portions consumed. Furthermore, our approach allows us to provide a consistent method for supporting various kinds of food. In this work, we focus on a particular subset of shapes for volume estimation including a cylinder and generalized extruded solids with flat tops and bottoms (e.g., bread slice).

3.1 False Information Removal

Automatic image segmentation generally includes undesired noise, such as holes, gaps, and bulges, or else the segmented may exclude necessary features. Such segmentation errors would cause errors in our feature extraction because noise along the boundaries of a segmented region would deform the original shape of a food. Since our feature extraction algorithm depends on the quality of the segmented image, we improve the segmented region of a food by removing false information such as noise on the boundaries. To filter the noise, we use mathematical binary morphology. Hence, from the segmented image (as shown in Figure 2, the middle image in each group) we generate a smoothed version of the segmented images (as shown in Figure 2, the right image in each group) note now unnecessary gaps and perturbations are removed.

3.2 Feature Extraction

Given a well-qualified segmented image, our algorithm extracts concrete features to size a food template shape. Although these features depend on template shapes, in our method, they contain three feature points and an average width obtained from the segmented image as well as interior edges from the original food image. These detected features determine the actual size of geometric elements to compute an estimated food volume.

Feature Points Extraction—To compute the shape of a food, we need at least three feature points: two points to estimate the width of the bottom area and one point to find a

height, as shown in Figure 3 (right) (see the three red dots). To find these points, we extract the contour of a segmented region and trace points ($p_i = (x_i, y_i)$) on the contour lines in order to detect the three points with high curvature. In Equation 1, the curvature C_i on the point p_i is computed from the difference between two standard deviations, S_{x_i} and S_{y_i}, on the point p_i. To control the effect of change in the x coordinate, we used a weight value, λ, depending on a template shape. For the result in Figure 3 (right), we used $\lambda = 1{:}5$. When the curvature C_i is higher than $1{:}0$, we select the point p_i as a feature point.

$$C_i = S_{y_i} - \lambda S_{x_i}$$
$$= \sqrt{\frac{1}{K}\sum_{k=1}^{K}(y_i{}^k - \overline{y_i}^1)^2} - \lambda \sqrt{\frac{1}{K}\sum_{k=1}^{K}(x_i{}^k - \overline{x_i}^1)^2}$$

(1)

where, $\overline{x_i}^1$ is an average of following K points starting from the point p_i in the x coordinate as shown in Equation 2. We used $K = 50$ for all results shown in this paper.

$$\overline{x_i}^1 = \frac{1}{K}\sum_{k=1}^{K}x_i{}^k$$

(2)

Global Geometric Feature Extraction—Since the boundaries of a segmented image are influenced by reflection and shadow from the picture-taking environment (e.g., lighting condition), global geometric properties of the segmented region is essential for our food volume estimation. The medial axis (also called the symmetric axis)[28] has been frequently used to represent and describe the global shape features.[24–26]

The medial axis of a shape is formally defined as at least two skeleton points that are equidistant from the the shape boundary. The use of the medial axis helps us to minimize volume estimation errors caused inaccuracies in detecting feature points. To this end, we first distinguish a central axis from the whole topology of the medial axis and use the central axis for extracting additional feature points (Figure 4 (middle right)). Next, we find point pairs on the boundary of the object that are on either side of the central axis (Figure 4 (rightmost)). The average length of these point pairs determine a potential width (radius for a cylindrical shape) of our template shape.

Interior Edge Detection—The analysis of the shape boundary is not enough in some cases where imperative geometric properties could be found within the interior area of a shape. Figure 5 shows an example of a bread slice with distinct interior edges. Here, our goal is to differentiate the side from the top surface such that the height from the bottom rim to the top surface can be computed. Moreover, the segmentation of the top surface provides us with more a accurate food shape when compared to an approximation using an ellipse. Thus, the volume of the bread slice is computed by multiplying these two measurements assuming that the shape of the bottom surface is the same to the top surface. As such, interior edge detection is utilized for specific foods with fairly flat and identical outlines of the top and bottom surfaces and the side distinguished by color. These cases would contain most sliced breads, such as a piece of toast, a garlic bread slice or a baguette slice.

For such interior edge detection and top surface segmentation, we employ the active contour methodology.[29] In computer vision, the active contour method has been used for in a wide range of problems including segmentation and edge detection.[30–33] The idea in the active contour method (or snakes) is, starting with a contour initialized near the desired object, the algorithm allows the contour (or snake) to deform so as to optimize the combination of

internal and external energy. Internal energy encourages an elastic and smooth shape, whereas external energy is based on strong edge attraction. In the classical edge based active contour models,[29, 34] the snakes depend on the gradient of an image to stop evolving contour on the boundary of an object. Our implementation, however, utilizes the region based active contour technique by Lankton and Tannenbaum[35] which is more robust with respect to initial contour placement and segmentation accuracy.

Given a food image (Figure 5 (leftmost)) and a corresponding segmented image (Figure 5 (middle left)), our algorithm initially cuts off the minimal rectangular region containing a food from the food image to reduce the computational overhead. It then places the initial snake on the top surface of a food to iteratively evolve the snake until the contour is on the boundary of the top surface (Figure 5 (middle right)). Therefore, we obtain interior edges and the top surface segmentation of a food. For calculating the height, as shown in Figure 5 (rightmost), we use two contour segments; the top segment (red line) along the interior edges and the bottom segment (white line) generated by translating the top contour segment to the bottom boundary of the segmented image. This provides an distance of the point pairs; one from the top segment and one the bottom segment.

4. RESULTS

To verify the accuracy of our volume estimation algorithm, we performed validation experiments using images of beverages (milk and orange juice in transparent container) and bread slices taken with the Apple iPhone (3GS) and Canon Powershot SD1100. For beverage images, we used automatically segmented images, and manually segmented images for bread images. The results of the beverage estimation indicated an improved accuracy of approximately 30% improved accuracy when removing false information in the segmented image using binary morphology. Further, using medial axis resulted in 10~ 20% improvement when compared to the use of only noise reduced images. Figure 6 shows examples of the computational results of our volume estimation algorithm, their real measurements with a beaker and used the segmented images. Table 1 shows the differences and ratios of the estimated volumes to the real measurements of 17 beverage images. The average relative error and standard deviation were about 11% and 8, respectively. We also compare the differences between estimated and measured volumes; 8 (N = 1 to 8) were underestimated and 9 (N = 9 to 17) overestimated. For bread slices as shown in Table 2, the results were 8% overestimated. Figure 7 shows some examples of our 3D reconstructed shape. We also found that our volume estimation method is influenced by the camera angle (viewing angle). In our experiments, the range of the appropriate camera angle is 30~45 degrees.

5. CONCLUSION AND FUTUREWORK

In this paper, we have presented a novel food volume estimation method from a single image by introducing food specific shape templates to deal with the complex and various food shapes and overcome insufficient information for 3D food volume reconstruction. In our approach, a novel feature point extraction algorithm was proposed including binary mathematical morphology to improve the accuracy of our algorithm by reducing noises in the segmented image. We also demonstrated how medial axis was used to extract geometrical features. Further, for volume estimation of free-shaped foods with the flat top and bottom surfaces, we applied the active contour method to determine interior edges and the top surface. Finally, we performed experiments to evaluate our approach using beverage images and achieved reasonable results (11% error). For generalized extruded solids seen as bread slices, promising computational volume results (8% error) was indicated. As future work, we plan to extend the range of shape templates to provide volume estimation for

various foods. Since accurate volume estimation for arbitrary shapes from a single image is still a computationally intractable problem, we are also planning to investigate a user-assisted method to improve accuracy and get a computationally feasible solution.

Acknowledgments

REFERENCES

1. Ogden CL, Flegal KM, Carroll MD, Johnson CL. Prevalence and trends in overweight among us children and adolescents, 1999–2000. JAMA: The Journal of the American Medical Association. 2002 Oct..288:1728–1732. [PubMed: 12365956]

2. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the united states, 1999–2004. JAMA: the journal of the American Medical Association. 2006 Apr..295:1549–1555. [PubMed: 16595758]

3. Rockville, M. The Surgeon General's call to action to prevent and decrease obesity. Washington DC: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General; 2001.

4. Taren D, Dwyer J, Freedman L, Solomons NW. Dietary assessment methods: where do we go from here? Public Health Nutrient. 2002:1001–1003.

5. Joy, Ngo; Anouk, Engelen; M. M. J. R. P. G.-S; Serra-Majem, L. A review of the use of information and communication technologies for dietary assessment. British Journal of Nutrition. 2009; 101:S102–S112. [PubMed: 19594959]

6. Lassen A, Poulsen S, Ernst L, Andersen K, Biltoft-Jensen A, Tetens I. Evaluation of a digital method to assess evening meal intake in a free-living adult population. Food and Nutrition Research. 2010; 1(0)

7. Glanz K, Murphy S, Moylan J, Evensen D, Curb JD. Improving dietary self-monitoring and adherence with hand-held computers: A pilot study. American Journal of Health Promotion. 2006

8. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. European Journal of Clinical Nutrition. 2009; 63:S50–S57. [PubMed: 19190645]

9. Yang, L.; Zheng, N.; Cheng, H.; Fernstrom, JD.; Sun, M.; Yang, J. Automatic dietary assessment from fast food categorization; Proceedings of the IEEE 34th Annual Northeast Bioengineering Conference; 2008.

10. Zhu, F.; Mariappan, A.; Boushey, CJ.; Kerr, D.; Lutes, KD.; Ebert, DS.; Delp, EJ. Technology-assisted dietary assessment; Proceedings of the IS&T/SPIE Conference on Computational Imaging VI 6814(1), 681411, SPIE; 2008.

11. Woo, I.; Otsmo, K.; Kim, S.; Ebert, DS.; Delp, EJ.; Boushey, CJ. Computational Imaging VIII 7533(1), 75330O, SPIE. 2010. Automatic portion estimation and visual refinement in mobile dietary assessment.

12. Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 2002; 2(4):509–522.

13. Zhu, Q.; Wang, L.; Wu, Y.; Shi, J. Contour context selection for object detection: A set-to-set contour matching approach; Proceedings of the European Conference on Computer Vision; 2008 October.

14. Berg, AC.; Malik, J. Geometric blur for template matching; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2001 December. p. 607-614.

15. Gavrila D. A bayesian exemplar-based approach to hierarchical shape matching. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 2007; 29:1408–1421.

16. Pavlidis T. A review of algorithms for shape analysis. Computer Graphics and Image Processing. 1978; 7(2):243–258.

17. Serra, J. Image Analysis and Mathematical Morphology. Orlando, FL, USA: Academic Press, Inc.; 1983.

18. Vincent L. Current topics in applied morphological image analysis. Current Trends in Stochastic Geometry and its Applications. 1997

19. Jeyalakshmi T, Ramar K. A modified method for speckle noise removal in ultrasound medical images. International Journal of Computer and Electrical Engineering. 2010 February.2

20. Schulze, MA.; Wu, QX. Noise reduction in synthetic aperture radar imagery using a morphology-based non-linear filter; Proceedings of Digital Image Computing: Techniques and Applications, Conference of the Australian Pattern Recognition Society; 1995. p. 661-666.

21. Heijmans HJAM. Self-dual morphological operators and filters. Journal of Mathematical Imaging and Vision. 1996; 6:15–36.

22. Peters RA, I. A new algorithm for image noise reduction using mathematical morphology. IEEE Transactions on Image Processing. 1995 May.4:554–568. [PubMed: 18290006]

23. Yang, G-Q.; Jiang, L-H.; Li, Y. Application of rough sets in binary morphology; Machine Learning and Cybernetics, 2006 International Conference on; 2006. p. 3446-3449.

24. Siddiqi K, Shokoufandeh A, Dickinson SJ, Zucker SW. Shock graphs and shape matching. International Journal of Computer Vision. 1999; 35:13–32.

25. Sundar H, Silver D, Gagvani N, Dickinson S. Skeleton based shape matching and retrieval. Shape Modeling International, 2003. 2003:130–139.

26. Pizer SM, Oliver WR, Bloomberg SH. Hierarchical shape description via the multiresolution symmetric axis transform. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 1987; 9(4):505–511.

27. Ho S-B, Dyer CR. Shape smoothing using medial axis properties. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 1986; 8(4):512–520.

28. Blum H. Biological shape and visual science (part i). Journal of Theoretical Biology. 1973; 38(2):205–287. [PubMed: 4689997]

29. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. International Journal of Computer Vision. 1988; 1:321–331.

30. Blake A, Isard M. Active Contours: The Application of Techniques from Graphics,Vision,Control Theory and Statistics to Visual Tracking of Shapes in Motion. 1998

31. Paragios, N.; Chen, Y.; Faugeras, O. Handbook of Mathematical Models in Computer Vision. Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2005.

32. Leymarie F, Levine MD. Tracking deformable objects in the plane using an active contour model. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 1993 June.15:617–634.

33. Chan T, Vese L. Active contours without edges. IEEE Transactions on Image Processing. 2001 Feb..10:266–277. [PubMed: 18249617]

34. Caselles V, Catte F, Coll T, Dibos F. A geometric model for active contours in image processing. Numerische Mathematik. 1993; 66:1–31.

35. Lankton S, Tannenbaum A. Localizing region-based active contours. IEEE Transactions on Image Processing. 2008; 17(11):2029–2039. [PubMed: 18854247]
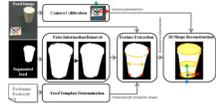
Figure 1.
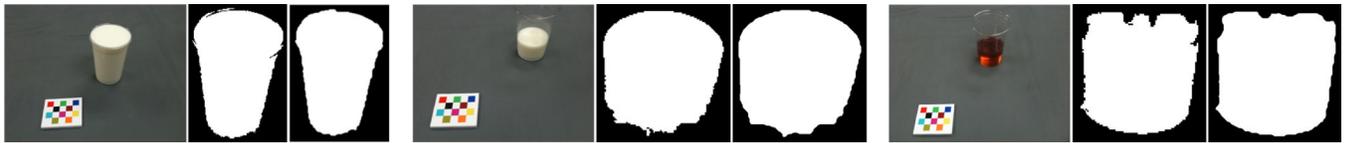Overview of our food volume estimation using food specific shape templates.

Figure 2.
Our false information removal using binary morphology. When considering every three images as a group, from left to right in each group, an original food image, a segmented image and our smoothed result of the segmented image. Each group shows full poured milk, half poured milk and half poured apple juice from the left to the right.
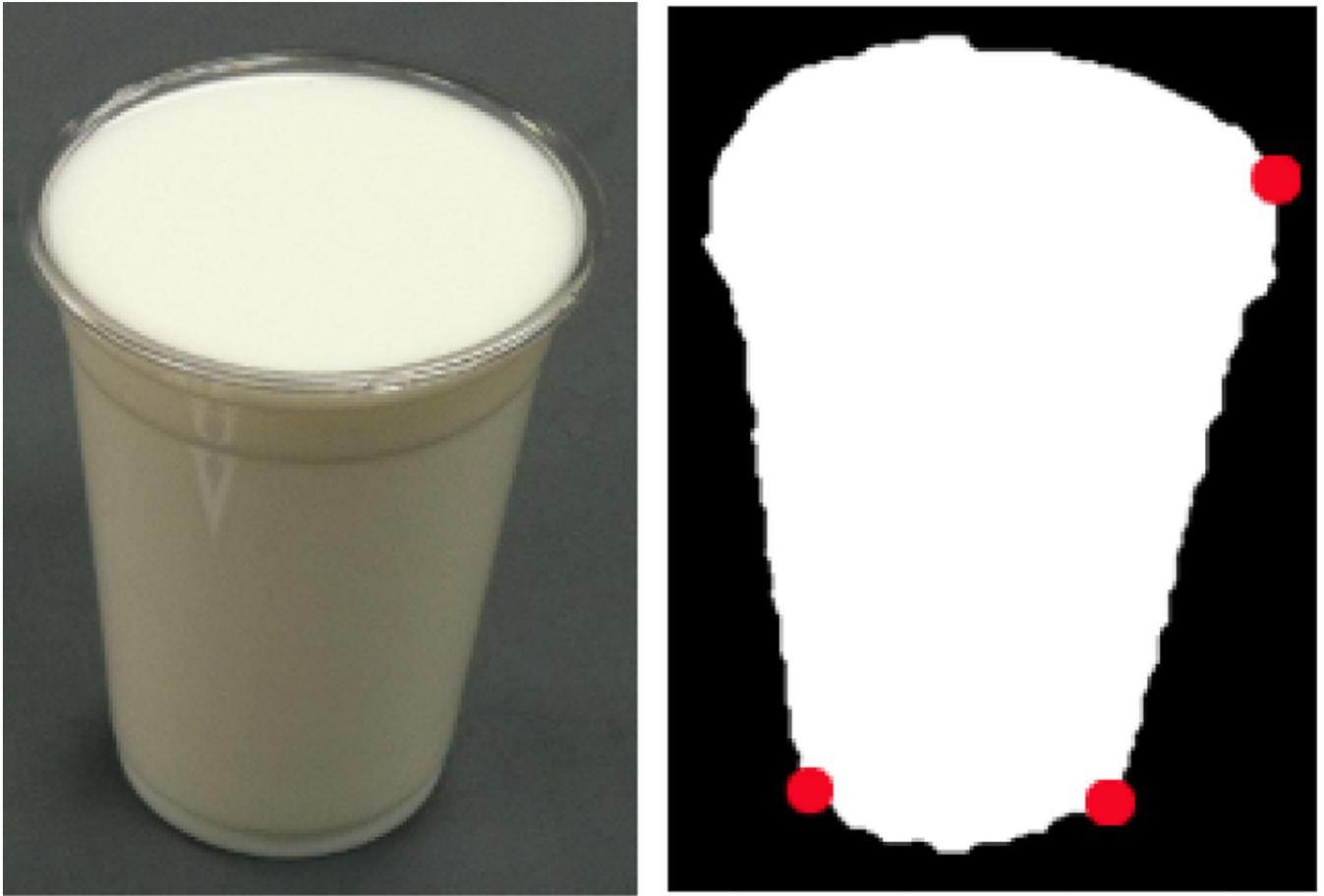
Figure 3.
(Left) an original food image and (right) three feature points.

Figure 4.
Geometric feature extraction using the medial axis. (Leftmost) a food image, (middle left) inaccurate feature point extraction, (middle right) medial axis generation and (rightmost) an average width of a shape along with the central axis of the medial axis.

Figure 5.
Feature extraction using the active contour. (Leftmost) a food image, (middle left) a segmented image, (middle right) active contour and (rightmost) height estimation using segments from active contour.

Figure 6.
Computational results of our volume estimation along with their real measurement and segmentation image.
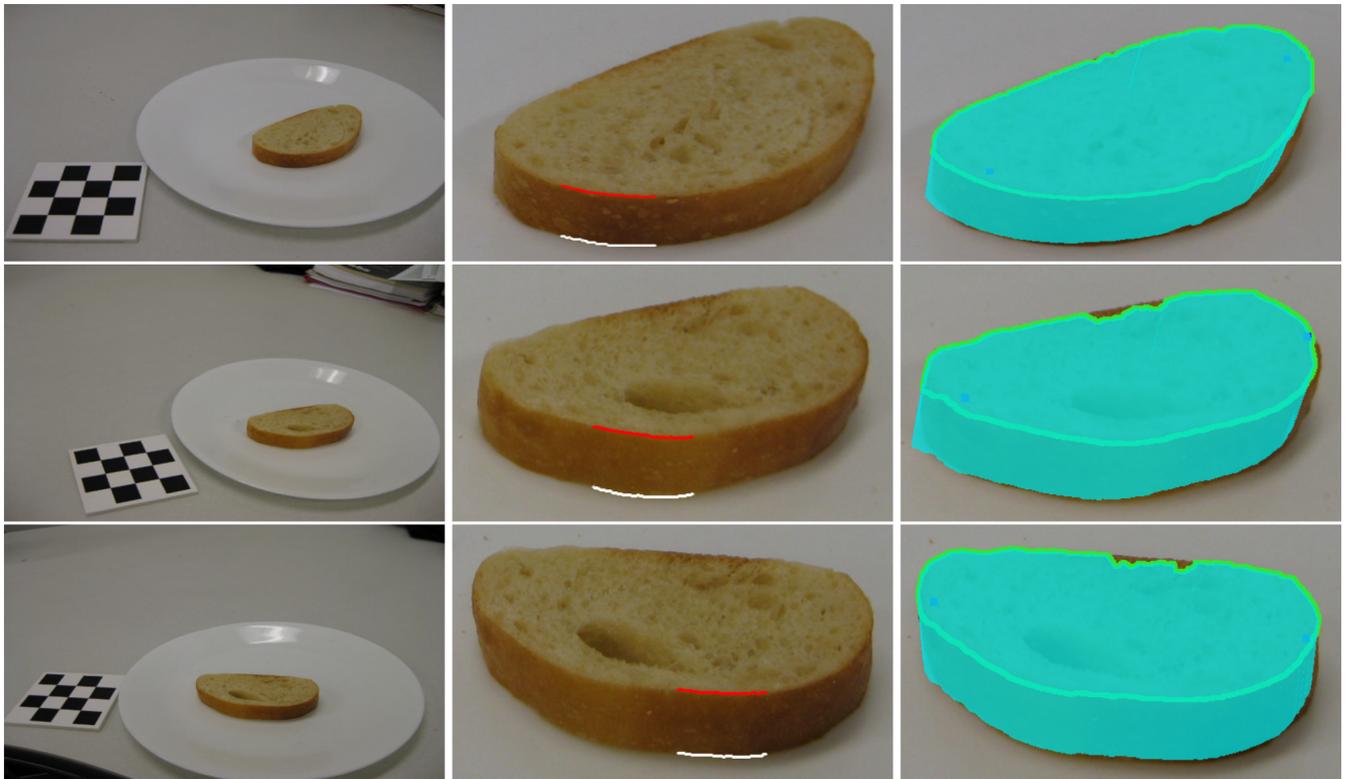
Figure 7.
Results of 3D volume reconstruction for different bread slice images. From left to right, food images, interior edge detection and 3D volumes reconstructed.

Table 1

Volume estimation results for beverages.

N	Measured volume	Estimated volume	Difference	Ratio of estimated volume to real measurement
1	137	112.13	−24.87	0.82
2	205	169.15	−35.85	0.83
3	137	120.58	−16.42	0.88
4	135	121.2	−13.8	0.9
5	205	183.85	−21.15	0.9
6	220	206	−14	0.94
7	220	212.86	−7.14	0.97
8	135	131.15	−3.85	0.97
9	70	71.46	1.46	1.02
10	220	226.07	6.07	1.03
11	137	143.78	6.78	1.05
12	137	146.14	9.14	1.07
13	205	229.61	24.61	1.12
14	65	73.61	8.61	1.13
15	70	81.98	11.98	1.17
16	65	77.14	12.14	1.19
17	65	85.8	20.8	1.32

Table 2

Volume estimation results for bread slices.

N	Top area (sq in)	Height (in)	Volume (cubic in)	Ratio of estimated volume to real measurement
1	6.71	0.80	5.37	1.25
2	6.74	0.81	5.46	1.27
3	6.62	0.77	5.10	1.19
4	6.00	0.78	4.68	1.00
5	6.19	0.84	5.20	1.11
6	6.33	0.81	5.13	1.09
7	5.42	0.80	4.34	0.99
8	5.50	0.78	4.29	0.98
9	5.48	0.70	3.84	0.88