SNAKES ASSISTED FOOD IMAGE SEGMENTATION

Y. He #1, N. Khanna #2, C.J. Boushey *3, E.J. Delp #4

School of Electrical and Computer Engineering, Purdue University
465 Northwestern Avenue, West Lafayette, IN 47907-2035, USA
1 he16@purdue.edu
2 khannan@ecn.purdue.edu
4 ace@ecn.purdue.edu

* *Cancer Epidemiology Program, University of Hawaii Cancer Center*
1236 Lauhala Street, Suite 407G, Honolulu, HI 96813, USA
3 cjboushey@cc.hawaii.edu

***Abstract*—In this paper we describe an image segmentation method for segmenting food items in images used for dietary assessment. Dietary assessment methods used to determine the foods and beverages consumed at a meal are essential for understanding the link between diet and health. Snakes, or active contours, are used extensively to locate object boundaries and segment images. Experimental results using classical snakes on food images show the problems associated with contour initialization and poor detection performance for food images. In this paper, we explore various methods of contour initialization and integrate a background removal method to improve the performance of food image segmentation. We describe the details of the proposed food image segmentation method and also evaluate our segmentation approach on food images.**

I. INTRODUCTION

Dietary assessment, the process of determining what someone eats during the course of a day, provides valuable insight for mounting intervention programs for the prevention of many chronic diseases including obesity and diabetes[1], [2]. Traditional dietary assessment, comprised of written and orally reported methods, is time consuming and tedious [3]. An online food-logging system is presented in [4], which can distinguish food images from other images, analyze the food balance, and visualize the log. Mobile telephones with build-in digital cameras and network connectivity have been shown to provide unique mechanisms for improving the accuracy and reliability of dietary assessment [5]. Our team at Purdue University and the University of Hawaii is developing a mobile telephone based image analysis system to automatically estimate the food consumed at an eating occasion from food images acquired by mobile devices[6], [7]. An essential part of this system is automatic food identification and image-based food volume estimation. The result of food volume estimation is highly dependent on the segmentation accuracy for food regions [8]. An example of ideal food image segmentation is shown in Figure 1.

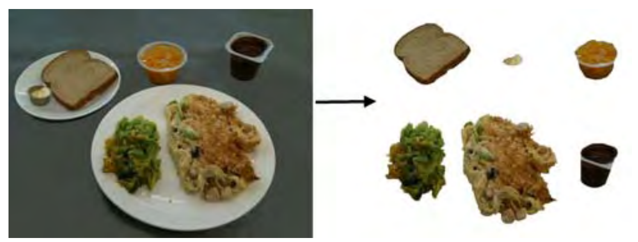


Fig. 1. Ideal Food Image Segmentation.

Snakes, or active contours, are used extensively in image processing and analysis applications, particularly for locating the object boundary [9]. A snake is an energy-minimizing spline guided by external constraints and influenced by the image structures that pull the spline toward features such as lines and edges [10]. Given an initial contour, active contour techniques guide the initial contour evolving toward the contour of the object of interest through minimizing a energy function. There are two general types of active contour models: *boundary-based active contours* [10], [11] and *region-based active contours* [12], [13]. For boundary-based active contours, the object is characterized by properties of its contour only. For region-based active contours, both internal and external regions of active contours are considered. Region dependency is also taken into account [14]. To accurately segment food regions, we consider region-based active contours in our implementation. In this paper, we show that the segmentation results of food images are highly dependent on the initial contours. One problem associated with using snakes on food images is the contours near non-food items (e.g. plates), which causes difficulties in food feature training and reduces the accuracy of food identification. Another problem is that some food items, especially the food items inside plates and other utensils, are often not detected. To address these problems, we explore various methods of contour initialization to reduce the correlation between initial contours and final segmentation results. We also integrate a background removal method with snakes to segment food images. A collection of food images acquired by mobile telephone cameras are tested in our experiments. We evaluated the segmentation results of food images by estimating the *precision* and *recall* for segmented regions.

II. Snakes Assisted Food Image Segmentation

Snakes, or active contours, are used for image segmentation by iteratively minimizing the segmentation energy. The basic idea is, starting with an initial contour, to deform the contour to the boundary of objects of interest, under some constrains from the image. In our implementation of active contours, two kinds of energies are considered: internal energy and external energy. The segmentation energy function is designed to reach a unique global minimum when the contours are the same as the edges of the objects we want to segment. Thus, we define the internal energy function and the external energy function as Equation 1 and Equation 2, respectively.

$$\begin{aligned} E_{int}(\Gamma) &= \int_{\Omega} \phi(\Gamma, x)\, \mathrm{d}x \\ &= \int_{\Omega} (f(x) - \mu(\Gamma))^2\, \mathrm{d}x \end{aligned} \tag{1}$$

$$\begin{aligned} E_{ext}(\Gamma) &= \int_{\Omega^c} \phi^c(\Gamma, x)\, \mathrm{d}x \\ &= \int_{\Omega^c} (f(x) - \mu^c(\Gamma))^2\, \mathrm{d}x \end{aligned} \tag{2}$$

where Γ is an oriented contour; $E_{int}(\Gamma)$ and $E_{ext}(\Gamma)$ denote the internal energy and the external energy, respectively. Ω is the internal set of the corresponding contour; Ω^c is the complement of Ω in the image domain which corresponds to the external set of the contour. $\phi(\Gamma, x)$ is sometimes referred to as the descriptor of the object of interest [15]. Similarly, $\phi^c(\Gamma, x)$ is the descriptor of the background. We define the descriptors in Equation 1 and Equation 2 as the mean square difference between the luminance of the current pixel and the average luminance of the internal (or external) area of the current contour Γ. $f(x)$ is the luminance value of pixel x; $\mu(\Gamma)$ and $\mu^c(\Gamma)$ are the average luminance value of pixels in the internal area Ω and the external area Ω^c, respectively. We can see from Equation 1 that if the object of interest is an ideal object with constant intensity, then $\phi(\Gamma, x)$ is equal to 0. The minimal energy is achieved when the final contour matches the edge of the object of interest. In our implementation, we design the final energy function as the sum of the internal energy and the external energy (Equation 3).

$$E(\Gamma) = E_{int}(\Gamma) + E_{ext}(\Gamma); \tag{3}$$

After defining the energy function, we still need to address the problem of contour initialization. The relative location of the initial contour and the objects of interest is essential to the final segmentation results. At first, we initialize the contour as a rectangular contour near the boundary of the image or in the middle of the image. We tested this initialization method on several food images with different food items. Some of these experimental results are shown in Figure 2 and indicate low segmentation accuracy. Figure 2 shows the original food images with the initial contours in red and the final segmentation results in blue.

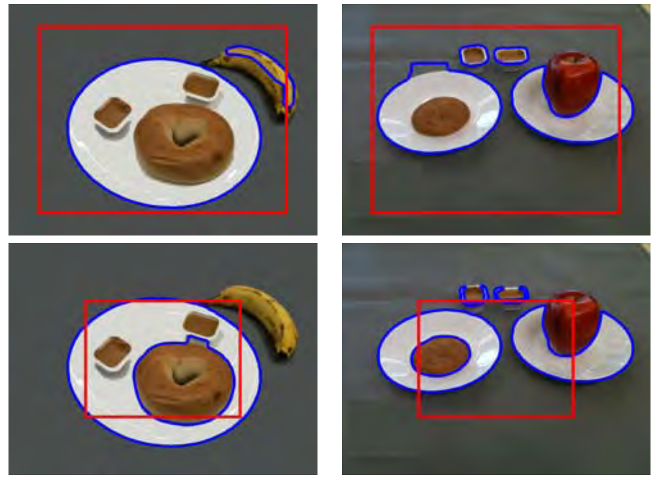


Fig. 2. Segmentation results for a rectangular initial contour. The initial contour is shown in red and the final contours of segmentation are shown in blue.

From the results in Figures 2, it is obvious that the single contour initialization method for food image segmentation is not robust since food images usually contain multiple food items. The segmentation result is highly dependent on the relative position of food items and the initial contours. Thus we need to use multiple initial contours. In the following experiments, we use multiple circles which are distributed evenly on the food images as initial contours. In Figure 3 the segmentation results for a different number of initial contours are presented. Improved segmentation is observed by comparing Figure 3 with Figures 2. We can also see from Figure 3 that some food items are still not detected. The plates or bowls are not counted as food items, thus the contours surrounding plates (or bowls) are undesirable contours, which introduce noise in food image segmentation. We improve the segmentation results by integrating a background removal method to remove the background regions before image segmentation. Our proposed method is illustrated in Figure 4.

To simplify the problem of food image segmentation, a tablecloth with uniform color are used in our dietary assessment studies. The tablecloth with uniform color is the most frequently occurring color in the scene. Therefore, we use color histogram for each color channel in the RGB color space to detect the most frequently occurring color. In our implementation, we use 32 bins for the color histogram, thus the number of pixels that fall into the i^{th} bin are:

$$h_{c\in(r,g,b)}(i) = \sum_{n=1}^{N} \delta(\lfloor \frac{x_n(c)}{8} + 1 \rfloor - i) \tag{4}$$

where $c \in (r, g, b), i \in [1, 32]$; $h_c(i)$ is the number of pixels whose c component falls into the i^{th} bin; N is the total number

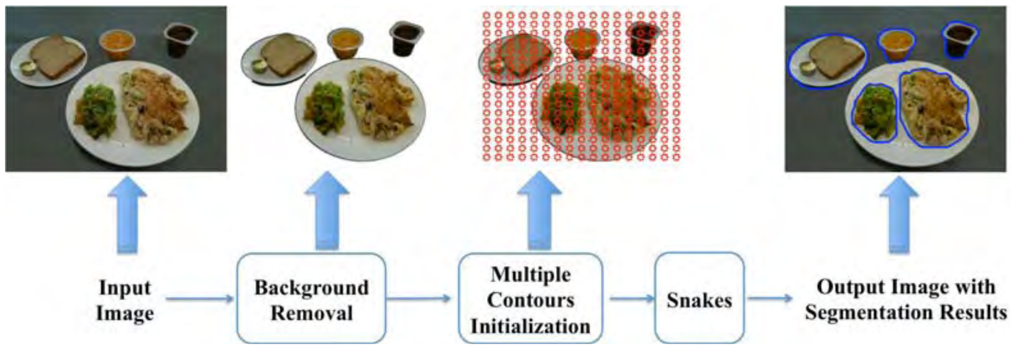
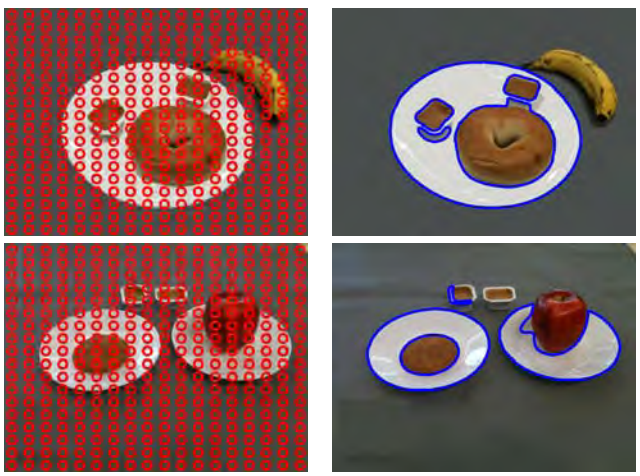


Fig. 4. Block diagram of our proposed method for food image segmentation.



(a) (b)

Fig. 3. Illustration of segmentation results for multiple initial contours distributed evenly: (a) the original food images with the location of the initial contours (in red); (b) the segmentation results in blue.

of pixels in the image; $x_n(c)$ is the value of the c component of pixel x_n; $\delta(x)$ is the impulse function which equals to 1 at $x = 0$ and 0 anywhere else.

Suppose the most frequently occurring color detected from the color histogram using Equation 4 falls in the bin l, where l is a three dimensional vector containing the bin number with the most frequently occurring R, G, B values, an image without background area is generated:

$$I_1(x,y) = \begin{cases} C_{bkg} & \text{if } \lfloor \frac{I_{org}(x,y)}{8} + 1 \rfloor = l \\ I_{org}(x,y) & \text{otherwise} \end{cases} \quad (5)$$

where (x, y) represents 2D coordinates of the original image; $I_{org}(x, y)$ is the value of the original image at (x, y); C_{bkg} is the color value used to indicate the background areas of food images. In this paper, the background color is set to white ($C_{bkg} = [255, 255, 255]$).

We did two experiments for 60 food images. In the first experiment we use snakes with multiple initial contours to the original food images. In the second experiment we first remove the background of an original food image, described in Equation 5, and then use snakes to generate the segmentation with multiple initial contours. The experimental results are presented in Figure 5. Figure 5(a) shows the original food images; Figure 5(b) shows the initial contours on the original images; (c) shows the segmentation results of snakes using the original background; Figure 5(d) shows the segmentation results of snakes after background removal.

III. Experimental Results

In additional to visual comparison of ground-truth segmentations and segmentations produced by snakes, we also examined quantitative evaluation by generating precision/recall scores for a range of input parameters (number of initial contours between 9 to 289) of the snakes method. For benchmarking image segemntation [16], [17], two properties are usually desired: objectivity and generality [18]. Objectivity means that the test images should have a suitably defined ground-truth segmentation so that the segmentation evaluation can be conducted objectively. Generality means that the test images should have a large variety so that the evaluation results can be extended to general images. In this paper, we adopted the idea of objectivity and generality in the evaluation of our segmentation results. To obtain a benchmark, for each parameter we tested 60 food images acquired by mobile telephone cameras in our studies. We divided these 60 images into two sets depending on the uniformity of their background. "Set-1" contains 30 food images with uniform background (Figure 5(a) first row) and "Set-2" contains 30 food images with some non-uniform background structure (Figure 5(a) second row). Different food items are used in each food image to ensure the variety of food images tested in our study. Human segmentation of the food items is used as ground-truth segmentation for all food images.

In general terms, there are two approaches to measure the segmentation quality: comparing the overlap between regions in two segmentations of the same image; and evaluating the agreement between corresponding segmentation boundaries [19]. Since the segmentation output of our method is an image with labeled boundary pixels, we evaluate the segmentation results by comparing the boundaries from automatic segmentation with the boundaries from ground-truth segmentation.

(a) (b) (c) (d)

Fig. 5. Segmentation results of snakes with multiple initial contours distributed evenly: (a) the original food images (first row: set-1; second row: set-2); (b) the food images with initial contours (red); (c) the segmentation results with original background; (d) the segmentation results after background removal.

Due to the computationally intensive nature of our method, we did the segmentation on images that have been down-sampled by a factor of 4 from their original size in our image database. Since the ground-truth segmentations have the same size as the original images, we also down-sample the ground-truth segmentation by a factor of 4. In [20] Martin proposed the use of precision/recall curves based on the region boundaries to evaluate segmentation consistency. Given a segmentation result from our snakes-based method S_{snakes}, and the ground-truth segmentation $S_{ground-truth}$, precision is defined as the proportion of boundary pixels in the automatic segmentation that correspond to boundary pixels in the ground-truth; while recall is defined as the proportion of boundary pixels in the ground-truth that were successfully detected by the automatic segmentation. Precision is sensitive to over-segmentation, while recall is sensitive to under-segmentation. High values in both precision and recall scores can only be achieved when the boundaries match in both segmentations.

$$\text{precision} = \frac{\text{Matched}(S_{snakes}, S_{ground-truth})}{|S_{snakes}|} \quad (6)$$

$$\text{recall} = \frac{\text{Matched}(S_{ground-truth}, S_{snakes})}{|S_{ground-truth}|} \quad (7)$$

where $|S_{snakes}|$ is the total number of automatically detected boundary pixels in the food images; $|S_{ground-truth}|$ is the total number of boundary pixels in the ground-truth images. In our evaluation, we set the matching threshold to be 5, which means for two boundary pixels are "matched" if their spatial location distance is within 5 pixels. The size of the down-sampled images are 512 $\times$ 384.

In our experiments, the initial contours are set to be in the same size and distributed evenly on the food images as shown in Figure 5(b). Precision and recall scores are calculated for all the images in our study. The turing curves resulting from the average precision and recall scores are shown in Figure 6 and Figure 7, respectively. We see from these two figures that the segmentation result has been improved by background removal.

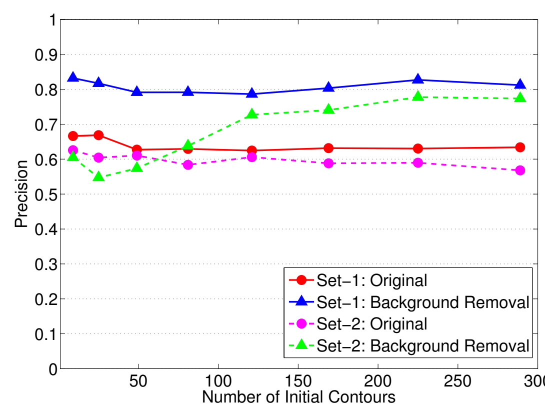


Fig. 6. Segmentation precision of food images with various numbers of initial contours.

IV. Conclusion

In this paper, we use active contours to segment food image and improve the segmentations by using multiple contour initialization and integrating a background removal model. The segmentation improvement has been shown by comparing our final results after background removal with the original segmentation results using snakes. We are currently investigating methods to improve and generalize the background model to better model typical eating occasions. We are also investigating how the segmentation approach described here improves our food classification [21] and volume estimation [8].

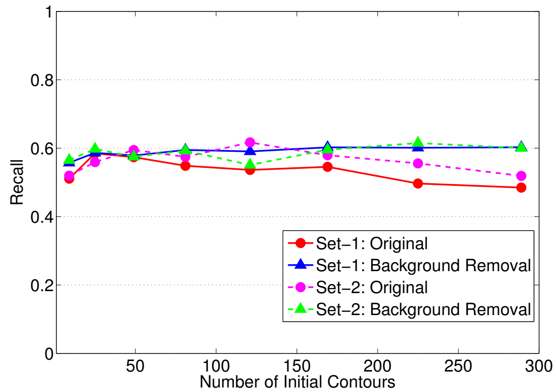


Fig. 7. Segmentation recall of food images with various numbers of initial contours.

ACKNOWLEDGMENT

This work was sponsored by grants from the National Institutes of Health under grants NIDDK 1R01DK073711-01A1 and NCI 1U01CA130784-01. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institutes of Health. Address all correspondence to Edward J. Delp, ace@ecn.purdue.edu or see www.tadaproject.org.

REFERENCES

[1] C. L. Ogden, M. D. Carroll, L. R. Curtin, M. M. Lamb, and K. M. Flegal, "Prevalence of high body mass index in US children and adolescents, 2007-2008," *Journal of the American Medical Association*, vol. 303, no. 3, pp. 242–249, January 2010.

[2] A. Fagot-Campagna, J.B. Saadinem, K.M. Flegal, and G.L.A. Beckles, "Diabetes, impaired fasting glucose, and elevated hba1c in US adolescents: the third national health and nutrition examination survey," *Diabetes Care*, vol. 24, no. 5, pp. 834–837, May 2001.

[3] F. E. Thompson, A. F. Subar, C. M. Loria, J. L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *Journal of the American Dietetic Association*, vol. 110, no. 1, pp. 48–51, January 2010.

[4] K. Kitamura, T. Yamasaki, and K. Aizawa, "Foodlog: Capture, analysis and retrieval of personal food images via web," *Proceedings of ACM multimedia workshop on Multimedia for cooking and eating activities*, pp. 23–30, November 2009.

[5] 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," *European Journal of Clinical Nutrition*, vol. 63 Suppl 1, pp. S50–57, February 2009.

[6] 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 Journal of Selected Topics in Signal Processing*, vol. 4, no. 4, pp. 756 –766, August 2010.

[7] M. Bosch, T. Schap, F. Zhu, N. Khanna, C.J. Boushey, and E.J. Delp, "Integrated database system for mobile dietary assessment and analysis," *Proceedings of the 1st IEEE International Conference Workshop on Multimedia Services and Technologies for E-health in conjunction with the International Conference on Multimedia and Expo*, Barcelona, Spain, July 2011, pp. 1 – 6.

[8] J. Chae, I. Woo, S. Kim, R. Maciejewski, F. Zhu, E.J. Delp, C.J. Boushey, and D.S. Ebert, "Volume estimation using food specific shape templates in mobile image-based dietary assessment," *Proceedings of the IS&T/SPIE Conference on Computational Imaging IX*, San Francisco, USA, February 2011, vol. 7873.

[9] C. Xu and J.L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359 –369, March 1998.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal Of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[11] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1995.

[12] R. Ronfard, "Region based strategies for active contour models," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 229–251, October 1994.

[13] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2029 –2039, November 2008.

[14] S. Jehan-Besson, M. Gastaud, F. Precioso, M. Barlaud, G. Aubert, and É. Debreuve, "From snakes to region-based active contours defined by region-dependent parameters," *Applied Optics*, vol. 43, no. 2, pp. 247–256, January 2004.

[15] É. Debreuve, M. Gastaud, M. Barlaud, and G. Aubert, "Using the shape gradient for active contour segmentation: from the continuous to the discrete formulation," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 1, pp. 47–66, May 2007.

[16] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *Proceedings of the IEEE International Conference on Computer Vision*, Vancouver, Canada, January 2001, vol. 2, pp. 416 – 423.

[17] F.J. Estrada and A.D. Jepson, "Quantitative evaluation of a novel image segmentation algorithm," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, June 2005, vol. 2, pp. 1132 – 1139.

[18] F. Ge, S. Wang, and T. Liu, "New benchmark for image segmentation evaluation," *Journal of Electronic Imaging*, vol. 16, no. 3, pp. 033011, July 2007.

[19] F. J. Estrada and A. D. Jepson, "Benchmarking image segmentation algorithms," *International Journal of Computer Vision*, vol. 85, no. 2, pp. 167–181, November 2009.

[20] D. R. Martin, *An Empirical Approach to Grouping and Segmentation*, Ph.D. thesis, Electrical Engineering and Computer Sciences Department, University of California, Berkeley, August 2003.

[21] M. Bosch, F. Zhu, N. Khanna, C.J. Boushey, and E.J. Delp, "Combining global and local features for food identification and dietary assessment," *Proceedings of the International Conference on Image Processing*, Brussels, Belgium, September 2011, pp. 1789 – 1792.