



Published in final edited form as:

ISSCS 2013 (2013). 2013 July ; 2013: .

Image Segmentation for Image-Based Dietary Assessment: A Comparative Study

Y. He^{*}, N. Khanna[†], C.J. Boushey[‡], and E.J. Delp^{*}

^{*}School of Electrical and Computer Engineering, Purdue University

[†]Department of Electronics and Communication Engineering, Graphic Era University

[‡]Cancer Epidemiology Program, University of Hawaii Cancer Center

Abstract

There is a health crisis in the US related to diet that is further exacerbated by our aging population and sedentary lifestyles. Six of the ten leading causes of death in the United States can be directly linked to diet. Dietary assessment, the process of determining what someone eats during the course of a day, is essential for understanding the link between diet and health. We are developing imaging based tools to automatically obtain accurate estimates of what foods a user consumes. Accurate food segmentation is essential for identifying food items and estimating food portion sizes. In this paper, we present a quantitative evaluation of automatic image segmentation methods for food image analysis used for dietary assessment. The experiments indicate that local variation is more suitable for food image segmentation in general dietary assessment studies where the food images acquired have complex background.

Index Terms

Image Segmentation; Active Contours; Local Variation; Normalized Cuts; Dietary Assessment

I. Introduction

Dietary assessment, the process of determining what someone eats during the course of a day, is essential for understanding the link between diet and health. 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 [1]. Our team at Purdue University and the University of Hawaii is developing a mobile telephone based image analysis system to estimate the food consumed at an eating occasion from food images acquired by the mobile telephone [2], [3]. Our goal is to automatically identify food items and estimate the volume of each food item. Food segmentation, segmenting food items from images acquired before and after each eating occasion, is an essential part of this system. The results of food identification and volume estimation are highly dependent on the segmentation accuracy [4], [5].

Various approaches are investigated in our system to segment food items in an image. In this paper we evaluate the food segmentation results of active contours [6], normalized cuts [7] and local variation [8] for food image segmentation. The quantitative evaluation of our methods is based on food images acquired during several diet studies conducted by the Department of Nutrition Science at Purdue University whereby participants were asked to take pictures of their food before and after each eating occasion [9]. The ground-truth for food image segmentation is obtained by human segmentation to evaluate various segmentation methods. The goal of this paper is to present an evaluation of the segmentation methods for food images. To obtain a meaningful comparison, we have tested the segmentation methods using many possible combinations of input parameters.

II. Segmentation Methods

Three segmentation methods are evaluated in this paper: active contours [6], normalized cuts [7], and local variation [8]. Before we proceed to the experimental evaluation part of this paper, it is convenient to briefly review the segmentation algorithms we have selected for comparison.

A. Active Contours

The basic idea of active contours is to deform an initial contour to the boundary of objects of interest by iteratively minimizing an energy cost function [6]. In our application of region-based active contours two types of energy functions are considered: internal energy (Equation 1) and external energy (Equation 2) [10], [11]:

$$E_{\text{int}}(\Gamma) = \int_{\Omega} \phi(\Gamma, x) dx = \int_{\Omega} (f(x) - \mu(\Gamma))^2 dx \quad (1)$$

$$E_{\text{ext}}(\Gamma) = \int_{\Omega^c} \phi^c(\Gamma, x) dx = \int_{\Omega^c} (f(x) - \mu^c(\Gamma))^2 dx \quad (2)$$

where Γ is an oriented contour; 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; $f(x)$ is the intensity of pixel x ; $\mu(\Gamma)$ and $\mu^c(\Gamma)$ are the average intensity of pixels in the internal area Ω and the external area Ω^c , respectively. $\phi(\Gamma, x)$ is sometimes referred to as the descriptor of the object of interest [12]. Similarly, $\phi^c(\Gamma, x)$ is the descriptor of the background.

The final energy function is defined as the sum of the internal energy and the external energy, which reaches a unique global minimum when the contours are same as the edges of the objects we want to segment. In active contours, the segmentation result is highly dependent on the relative position of food items and the initial contours. To reduce the correlation, we initialize the contours as multiple circles which are distributed evenly on the food image. The details of the initialization method is described in [13].

B. Normalized Cuts

Normalized Cuts is a graph based image segmentation method which treats the segmentation problem as the partition of a graph $G = (V, E, W)$ [7]. Each image pixel is assigned as a node of a graph (V). The weight of an edge (E) between image pixels, $W(i, j)$, is a measure of the affinity between two pixels i and j based on their intensity similarity and spatial proximity. The entire image is modeled as a weighted, undirected graph. The image is segmented into disjoint sets by minimizing the normalized cut measure:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \quad (3)$$

where $cut(A, B)$ is the cut cost of (A, B); and $assoc(A, V)$ is the total weight of connections between the region A and the full graph V , and measures the strength of the association between the region A and the entire image.

In this paper, we estimate the weight between two pixels using *intensity* and *intervening contours* proposed in [14]. The distance between pixels is also taken into account in the intensity function, because the intensity cue alone often gives poor segmentations [14]. The intervening contours function measures the magnitude of image edges between them. Let $X_A(i) = 1$ if and only if pixel i belongs to the segment A , and D be a diagonal matrix where $D(i, i) = \sum_j W(i, j)$. Then the segmentation measure could be formatted in a matrix form:

$$Ncut(A, B) = \frac{X_A^T (D - W) X_A}{X_A^T D X_A} + \frac{X_B^T (D - W) X_B}{X_B^T D X_B} \quad (4)$$

Calculating the optimal Ncut graph partition is a NP hard problem. However, an approximate solution is provided by computing the K eigenvectors corresponding to the K largest eigenvalues in:

$$(D - W)x = \lambda Dx \quad (5)$$

C. Local Variation

Similar to normalized cuts, local variation proposed in [8] is also a graph based image segmentation method. Weights on each edge measure the dissimilarity between pixels. This method segments the image based on the degree of variability in neighboring regions of the image. The internal difference of a segmented region is defined to be the largest weight in the minimum spanning tree, $MST(A, E)$, of the region:

$$Int(A) = \max_{e \in MST(A, E)} w(e) \quad (6)$$

where $w(e)$ is the weight of an edge e . The difference between two segmented regions is defined to be the minimum weight edge connecting the two regions:

$$\text{Dif}(A, B) = \min_{p \in A, q \in B, (p, q) \in E} w(p, q) \quad (7)$$

The proposed method segments two regions if the difference between the two regions $\text{Dif}(A, B)$ is large relative to the internal difference within at least one of the two regions. The degree to which the difference between regions must be larger than minimum internal difference is controlled by a threshold k :

$$\text{minInt}(A, B) = \min\left(\text{Int}(A) + \frac{k}{|A|}, \text{Int}(B) + \frac{k}{|B|}\right) \quad (8)$$

where $|A|$ denotes the size of A .

III. Experimental Results and Quantitative Evaluation

Examples of food segmentation results using different segmentation methods are presented in Figure 1. Figure 1 (a) shows the original food images. Figure 1 (b) shows the food images after background removal. We generate a foreground-background image by labeling the most frequently occurring colors as the background color. This simple image preprocessing step may help to reduce the noise in food images and improve the segmentation results. Figure 1 (c), (d) and (e) corresponds to the food segmentation results using active contours, normalized cuts and local variation respectively.

In our experiments, we investigated 120 food images from our dietary assessment studies. Human segmentation of the food items is used as ground-truth segmentation for all food images. We evaluate the segmentation methods by computing the *precision* and *recall* of segmentation regions with respect to human ground-truth segmentations [15]. Precision and recall are particularly attractive as measures of segmentation quality because they are not biased in favor of over-segmented or under-segmented images [16]. Precision measures the percentage of segmented regions in the automatic segmentation that correspond to ground-truth segmentation (Equation 9); while recall is defined as the percentage of segmented regions in ground-truth segmentation that are detected in automatic segmentation (Equation 10).

$$\text{precision} = \frac{\text{Matched}(S_{\text{ground-truth}}, S_{\text{autoseg}})}{|S_{\text{autoseg}}|} \quad (9)$$

$$\text{recall} = \frac{\text{Matched}(S_{\text{autoseg}}, S_{\text{ground-truth}})}{|S_{\text{ground-truth}}|} \quad (10)$$

where $|S_{ground-truth}|$ is the total number of food items in ground-truth segmentation; $|S_{autoseg}|$ is the total number of segmented regions in the food image using automatic segmentation methods. In our evaluation, we set the matching threshold to be 80%, which means for any two particular regions A, B , we say they are “matched” if $\text{Matched}(A, B)$ in Equation 11 equals to 1.

$$\text{Matched}(A, B) = \begin{cases} 1 & \text{if } \frac{|A \cap B|}{|A \cup B|} > 80\% \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We test each segmentation method for different input parameters. The input parameter for active contours is the number of initial contours. The input parameter for normalized cuts is the number of desired segmentations. Since there are rarely more than 8 food items per eating occasion, we tested the segmentation method for the number of desired segmentations in the range [6, 15]. Taking into account the background area and the plates or bowls, we set the desired number of segmentations larger than 8. The input parameter for local variation is k in Equation 8 that roughly controls the size of the segmented regions. The precision and recall of different segmentation methods are presented in Figure 2. We can see from our experimental results that recall is usually higher than precision. The reason is that only the food items are segmented from the images in ground-truth segmentation, but automatic segmentation includes other segmented areas such as table and plates.

IV. Conclusion

Food image segmentation is an important step in mobile dietary assessment. In this paper, we use precision and recall to evaluate three image segmentation methods, namely active contours, normalized cuts and local variation, for food image segmentation. Based on our experimental results, local variation is more stable to the changes of input parameters than the other two segmentation methods. As the number of initial contours increases, active contours can achieve similar precision and recall scores as local variation, but the computation complexity also increases. Considering both precision/recall and the computation complexity, we adopt the framework of local variation in our food image analysis system.

Acknowledgments

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.

References

1. 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 Feb; 63(Suppl 1):S50–S57. [PubMed: 19190645]
2. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, Delp EJ. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE Journal of Selected Topics in Signal Processing*. 2010 Aug; 4(4):756–766. [PubMed: 20862266]

3. Bosch, M., Schap, T., Zhu, F., Khanna, N., Boushey, CJ., Delp, EJ. 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: 2011 Jul. Integrated database system for mobile dietary assessment and analysis; p. 1-6.
4. Zhu, F., Bosch, M., Schap, T., Khanna, N., Ebert, DS., Boushey, CJ., Delp, EJ. Proceedings of the IS&T/SPIE Conference on Computational Imaging IX. Vol. 7873. San Francisco, USA: 2011 Jan. Segmentation assisted food classification for dietary assessment.
5. Chae, J., Woo, I., Kim, S., Maciejewski, R., Zhu, F., Delp, EJ., Boushey, CJ., Ebert, DS. Proceedings of the IS&T/SPIE Conference on Computational Imaging IX. Vol. 7873. San Francisco, USA: 2011 Feb. Volume estimation using food specific shape templates in mobile image-based dietary assessment.
6. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. International Journal of Computer Vision. 1988; 1(4):321–331.
7. Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000 Aug; 22(8):888–905.
8. Felzenszwalb, PF., Huttenlocher, DP. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: 1998 Jun. Image segmentation using local variation; p. 98-104.
9. Six BL, Schap TE, Zhu F, Mariappan A, Bosch M, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Evidence-based development of a mobile telephone food record. Journal of the American Dietetic Association. 2010 Jan.110:74–79. [PubMed: 20102830]
10. Ronfard R. Region based strategies for active contour models. International Journal of Computer Vision. 1994 Oct; 13(2):229–251.
11. Lankton S, Tannenbaum A. Localizing region-based active contours. IEEE Transactions on Image Processing. 2008 Nov; 17(11):2029–2039. [PubMed: 18854247]
12. Debreuve É, Gstaad M, Barlaud M, Aubert G. Using the shape gradient for active contour segmentation: from the continuous to the discrete formulation. Journal of Mathematical Imaging and Vision. 2007 May; 28(1):47–66.
13. He, Y., Khanna, N., Boushey, CJ., Delp, EJ. Proceedings of IEEE International Workshop on Multimedia Signal Processing. Banff, Canada: 2012 Sep. Snakes assisted food image segmentation; p. 181-185.
14. Cour, T., Benezit, F., Shi, J. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 2. Washington, DC, USA: 2005 Jun. Spectral segmentation with multiscale graph decomposition; p. 1124-1131.
15. Martin, DR. Ph.D. thesis. Berkeley: Electrical Engineering and Computer Sciences Department, University of California; 2003 Aug. An Empirical Approach to Grouping and Segmentation.
16. Estrada FJ, Jepson AD. Benchmarking image segmentation algorithms. International Journal of Computer Vision. 2009 Nov; 85(2):167–181.

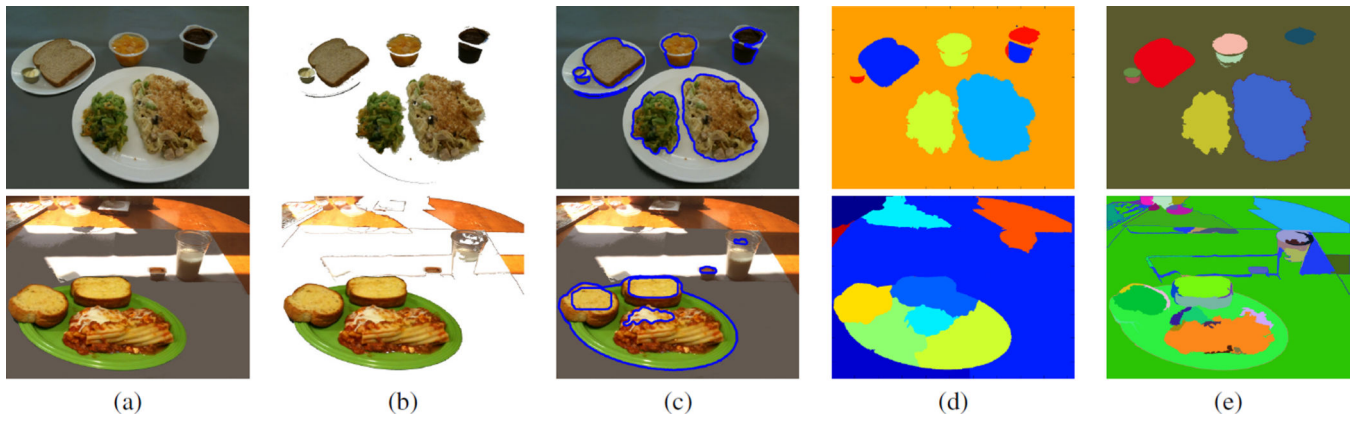


Fig. 1. Comparison of segmentation results using selected segmentation methods: (a) original food images; (b) food images after background removal; (c) segmentation results of active contours; (d) segmentation results of normalized cuts; (e) segmentation results of local variation.

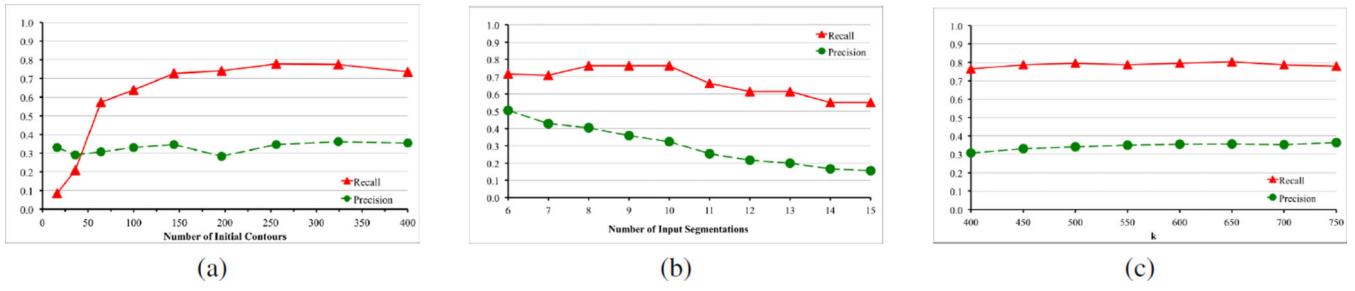


Fig. 2.
Precision and recall for (a) active contours, (b) normalized cuts and (c) local variation.