Goldilocks and the Three Parameters:Empirically Finding the "Just Right" for Segmenting Food Images for the AFINI-T System

Alexander MacLean[1] Audrey G. Chung[1]
Kaylen J Pfisterer[1,2] Devinder Kumar[1]
Robert Amelard[2] Alexander Wong[1,2]

[1]University of Waterloo, Waterloo, Canada
[2]Schlegel-UW Research Institute for Aging, Waterloo, Canada

Abstract

Measuring nutritional intake is a tool that is critical to the monitoring of health, both as an individual or of a group. It is especially important in the monitoring of those at risk for malnutrition, an issue which costs billions of dollars globally, and current methods used in practice are manual, time-consuming, and have inherent biases and inaccuracies. This study proposes a novel imaging system with a superpixel-based segmentation algorithm as part of an automated nutritional intake system. The study also examines three important parameters of the algorithm and their ideal values; region size and spatial regularization for superpixel segmentation, as well as spatial weighting in clustering. The experimental results demonstrate that the proposed system is effective in segmenting an image of a plate into its constituent foods.

1. Introduction

Measuring nutritional intake is vital for understanding and improving personal health, and is necessary to ensure that nutritional requirements are met [1] as malnutrition costs the health care industry and can lead to mortality [2]. This is especially important for vulnerable populations such as older adults living in long term care (LTC) homes; one in four older adults are at risk for malnutrition [3]. However, tracking intake for all LTC residents can be time-consuming and imprecise, which introduces errors into the system and often results in failure in tracking.

An automatic quantitative nutrition tracking system to measure food intake may provide a powerful solution to the "pen-and-paper" methods, which should allow a higher level of accuracy as well as reduce the time needed for tracking [4]. Fundamentally, such an imaging system must identify and separate food groups for tracking purposes. Such systems have been proposed, such as a segmentation system using Normalized Cuts based on intensity and colour of pixels [5]; however, this segmentation method included non-food items in the food classes, which led to inaccurate segmentation and would cause poor downstream nutritional analyses. In this study, a superpixel-based food group segmentation method was evaluated on simulated plates, with qualitative analyses being performed. To the best of the authors' knowledge, this is the first superpixel-based image segmentation method used for food and nutritional tracking, and is a first step for the proposed Automatic Food Imaging and Nutrition Intake Tracking (AFINI-T) System.

The methods section describes the methodology of the experiment and the steps involved in the algorithm. Following that, the results are discussed, with the effects of three keys parameters, region size and spatial regularization for superpixel segmentation and spatial weighting for clustering, being emphasized. The following discussion overviews limitations of the system and areas of future direction, and finally the conclusion discusses the impact of the experiment.

2. Methods

The proposed image processing method involves conversion from an image of a plate of food to a segmentation mapping which distinguishes between different classes of food on the plate. An overview of the algorithm, from the initial image to the final segmentation into the classes of food, is provided in Figure 1. The algorithm is composed of the following steps.

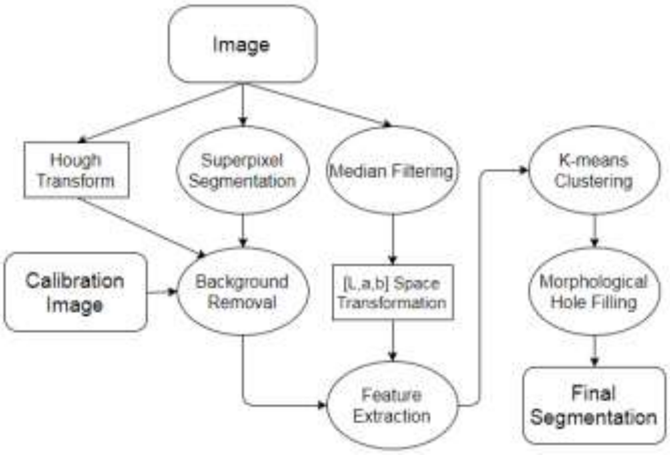

Figure 1: Overview of the food segmentation algorithm.

Superpixel segmentation. The image was segmented into smaller regions similar to a patch for the purpose of downstream clustering by simple iterative linear clustering (SLIC) superpixels [6]. The superpixel algorithm used k-means clustering to partition the image into regions which are more homogenous and consistent, based on tunable parameters for region size and spatial regularization. The algorithm returned labels for the region for each pixel of the image. Examples of the results of this superpixel segmentation with varying parameters are shown in Figure 3 and Figure 4.

Background removal. The desired output from this stage was to only pass superpixels mostly comprised of food into the clustering algorithm. First, a circular Hough Transform [7] was used to remove the background outside of the plate. Then, using a calibration image, the superpixels that contained >50% plate, based on image subtraction and spectral thresholding, were masked out of the image. This process is shown in Figure 2.

Figure 2: From left to right; a) the original image of the plate; b) the image after removing the background from a Hough Transform; c) the image after thresholding to remove the plate itself from the image.

Median filtering. A median filter was applied to the original image to increase the homogeneity of the components of the image, increasing the similarity of parts of the image that are in the same region and should be part of the same class, while maintaining the edges of objects without introducing new information into the image.

Feature extraction. A 4D spectral-space feature vector of [a,b,γx,γy] was then extracted for each superpixel, with [a,b] being the normalized mean values for the chroma channels of the pixels contained in each superpixel after a transformation to the L,a,b colour space of the median filtered image, and [x,y] representing the coordinates for the centroid of the superpixel. The L channel was omitted from clustering as to only consider chroma information. The parameter γ is a weight that was applied to the spatial information, and examples of varying the γ parameter are shown in Figure 5.

K-means clustering. K-means clustering [8] was performed using feature vectors for each superpixel to segment the image into classes representing each food, with the number of clusters set to the number of foods on the plate.

Morphological hole-filling. A morphological hole-filling approach [9] was used to remove small inconsistencies in the clustering, under the assumption that the classes of food were separable and none were contained in any others.

3. Results

Three important parameters (region size, regularization, γ) for the proposed spectral-spatial superpixel segmentation method were qualitatively evaluated to identify accuracy and precision of segmentation. For each parameter, three examples of parameter tuning are presented to illustrate the impact of small, large, and empirically optimized parameter settings on the final segmentation.

Figure 3 illustrates variation of the region size parameter for the superpixel algorithm. The larger the region size value given, the larger the area of each individual resulting superpixel. The region size parameter describes the initial size of the regions during superpixel segmentation, with higher values leading to larger regions with less homogeneity and lower values leading to smaller regions each with more homogeneity. Small values for the region size parameter (Figure 3a) led to failure to remove darker "islands" (e.g., marked plate labels), and the superpixels were not large enough to benefit from the ability of the algorithm to locate regions of homogeneity. Conversely, for large values for this parameter the outer pixels tended to contain both food and plate (e.g., the potatoes in Figure 3c, where the value of the parameter was 25). The result: the values taken for the feature vector from these pixels were skewed. With region size set to the empirically optimized value of 15 pixels as in Figure 3b, these issues were largely mitigated, as the pixels were large enough to be effective but also had a strong distinction between the food and plate, and ultimately led to improved clustering.

Figure 4 shows variation of the spatial regularization parameter; increasing the value forced the superpixel to resemble traditional equal-sized pixel patches with straighter edges. Conversely, reducing the value resulted in superpixels with increased more variability in superpixel shape and size. When the spatial regularizer was set to small values (e.g., 0.0001 in Figure 4a), the algorithm created superpixels based on inhomogeneity; these superpixels may take any shape and thus tend to over-separate foods into regions. The result: over-clustering within the same food item. Conversely, for large values (e.g., 1.0 in Figure 4c), too stringent of requirements on the shape of the superpixel were imposed, leading to edges which were too forced and more similar to pixel patches thus minimizing the benefit of using superpixels over pixel patches. As a result, too many pixels contained both food and background, resulting in inaccurate removal of some food and poorer clustering performance. The ideal value of 0.01 as shown in Figure 4b balanced these two pressures, creating superpixels that followed the natural borders of the image while not overfitting the differences within the foods.

Figure 5 shows variation of the γ parameter for weighting the spatial components of the feature vector during clustering. Spatial information was used since any two regions close to each other were more likely to belong to the same food class; however, the importance of this relative spatial information needed to be tuned in comparison to the importance of the spectral chroma information. Intuitively, regions in closer proximity to each other were more likely to be related to each other than more distant regions (i.e., two superpixels on adjacent edges of different foods were more closely related than two superpixels on opposite ends of the same food item). For very small values of the spatial weight parameter (i.e., weighing spatial information too low), as in Figure 5a where the value was 0.05, there was a greater risk of ignoring the relevance of spatial proximity. However, for very large values of spatial weighing (e.g., Figure 5c with a value of 2.0), spatial proximity dominated the clustering algorithm. Using the empirically found γ value of 0.4 appeared to provide the best middle ground, as shown in Figure 5b, as the spatial information helped to improve the homogeneity of the clustering while not being so large as to overpower the spectral information.

4. Discussion

The algorithm was developed specifically for images as shown in Figure 2a, where the food is located on a circular, white plate, and calibration images with an empty plate from the same experimental setup are available. Thus, while the superpixel segmentation and k-means clustering should function similarly under different conditions, the processing to remove the background of the image will need to be generalized to allow it to work in other varying environments. This, along with general improvements to the efficacy and precision of the background removal, is one of the primary directions of future research.

5. Conclusion

The results provide evidence that the algorithm is effective in segmenting an image of a plate into the food items present. Tuning the region size and spatial regularization parameters for superpixel segmentation and the spatial weighting parameter for feature extraction and clustering were critical for the development of the algorithm, and the ideal empirical values for the parameters are 15, 0.01, and 0.4 respectively. Future work will focus on refinement of this algorithm with attention to generalizability (e.g., for other plates and conditions) with fewer constraints.

6. Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs Program, and the Ontario Ministry of Research and Innovation.

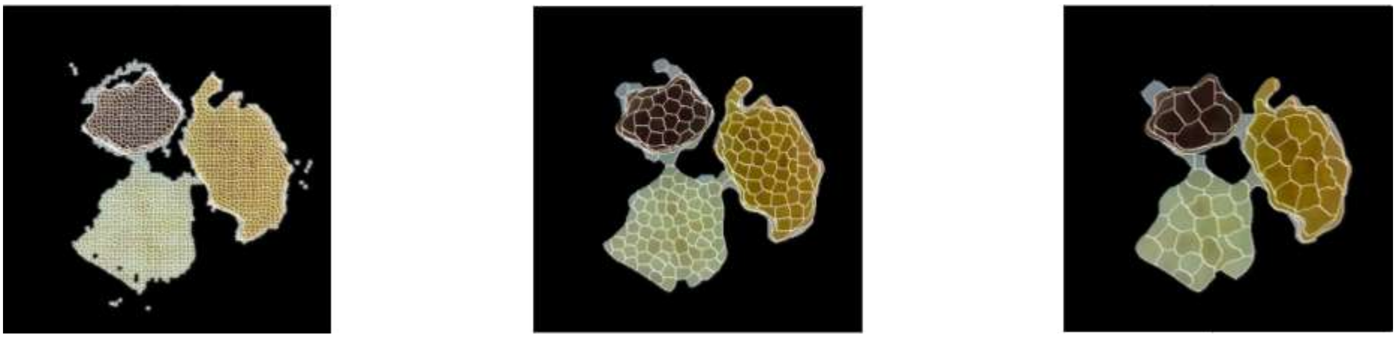

Figure 3: Results of superpixel segmentation, with variation of the region size parameter with a constant spatial regularization parameter of 0.01. From left to right, the values of the parameter are: a) 5, b) 15 (ideal), and c) 25.

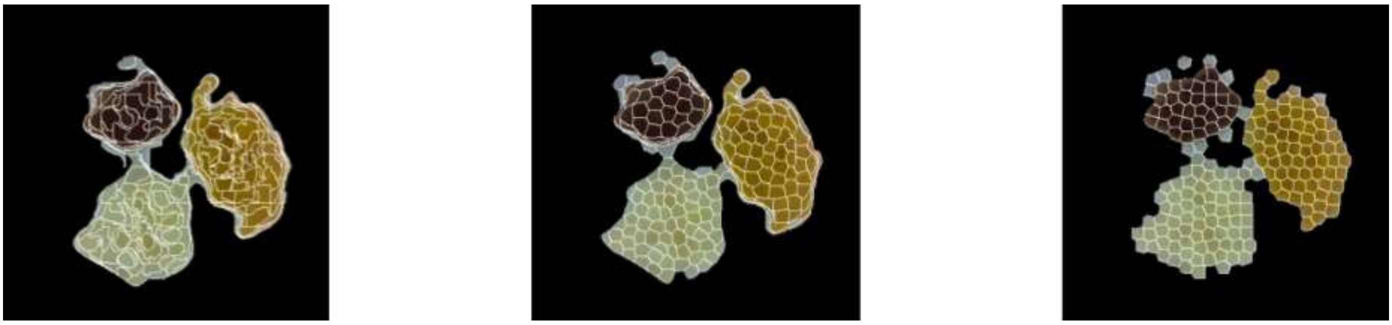

Figure 4: Results of superpixel segmentation, with variation of the spatial regularization parameter with a constant region size parameter of 15. From left to right, the values of the parameter are: a) 0.0001, b) 0.01 (ideal), and c) 1.

Figure 5: Results of k-means clustering and morphological hole-filling, with variation of the spatial weighting parameter for feature extraction. From left to right, the values of the parameter are: a) 0.05, b) 0.4 (ideal), and c) 2.0.

References

[1] L. Penn, H. Boeing, C. J. Boushey, L. O. Dragsted, J. Kaput, A. Scalbert, A. A. Welch, and J. C. Mathers. Assessment of dietary intake: Nugo symposium report. Genes & Nutrition, 5(3):205, 2010.

[2] M. Pirlich and H. Lochs. Nutrition in the elderly. Best Practice & Research Clinical Gastroenterology, 15(6):869–884, 2001.

[3] Kaiser, M. J., Bauer, J. M., Rämsch, C., Uter, W., Guigoz, Y., Cederholm, T., ... & Tsai, A. C. (2010). Frequency of malnutrition in older adults: a multinational perspective using the mini nutritional assessment. Journal of the American Geriatrics Society, 58(9), 1734-1738.

[4] D. A. Williamson, H. R. Allen, P. D. Martin, A. J. Alfonso, B. Gerald, and A. Hunt. Comparison of digital photography to weighed and visual estimation of portion sizes. Journal of the American Dietetic Association, 103(9):1139–1145, 2003.

[5] F. Zhu, M. Bosch, T. Schap, N. Khanna, D. Ebert, C. J. Boushey, and E. J. Delp. Segmentation Assisted Food Classification for Dietary Assessment. Proc. IS&T/SPIE Conf. Computat. Imag. IX,7873, 2011.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC superpixels compared to state-of-the-art ¨ superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2274–2282, 2012.

[7] T. J. Atherton and D. J. Kerbyson. Size invariant circle detection. Image and Vision Computing, 17(11):795–803, 1999.

[8] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In Proceedings of the 18th Annual ACMSIAM Symposium on Discrete Algorithms, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[9] P. Soille. Morphological Image Analysis: Principles and Applications. Springer Science & Business Media, 2013.