# FOOD IMAGE ANALYSIS: THE BIG DATA PROBLEM YOU CAN EAT!

*Yu Wang*[1], *Shaobo Fang*[1], *Chang Liu*[1], *Fengqing Zhu*[1], *Deborah A Kerr*[3], *Carol J Boushey*[2], *and Edward J Delp*[1]

[1] School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA
[2] Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA
[3] School of Public Health, Curtin University, Perth, Australia

## ABSTRACT

Six of the ten leading causes of death in the United States can be directly linked to diet. Measuring accurate dietary intake, the process of determining what someone eats is considered to be an open research problem in the nutrition and health fields. We are developing image-based tools in order to automatically obtain accurate estimates of what foods a user consumes. We have developed a novel food record application using the embedded camera in a mobile device. This paper describes the current status of food image analysis and overviews problems that still need to be addressed.

## 1. INTRODUCTION

Six of the ten leading causes of death in the United States, including cancer, diabetes, and heart disease, can be directly linked to diet. Measuring accurate dietary intake is considered to be an open research problem in the nutrition and health fields. Methods for dietary assessment and evaluation continue to be a challenge. Traditional dietary assessment, e.g. dietary record, is a time consuming and tedious process, which requires individuals to keep detailed written reports for 3-7 days of all food or drink consumed [1, 2]. Mobile telephones provide a unique mechanism for collecting dietary information and monitoring personal health. By February 2016, 72% of American adults were smartphone owners and there has been a noticeable rise in mobile phone and internet usage in the past few years in the emerging and developing nations [3]. The importance of using eating occasion images to record and estimate dietary intake versus classical handwritten food record approaches has been highlighted in our earlier work [4, 5].

For the past 9 years we have been investigating the use of images that a user takes of their meal before and after eating occasions to assess dietary intake. We have developed a system, known as the Technology Assisted Dietary Assessment System (TADA), to acquire and process food images [6, 4, 7]. The TADA system and the associated mobile Food Record (mFR) mobile application allows users to acquire food images using a mobile telephone (see Figure 1). Image processing and computer vision analysis methods are then used to determine the food type, volume, the energy (kilocalories) and nutrients of the food [8, 9, 7, 10]. The TADA system has been used for more than 14 scientifically implemented user studies, including environments in the wild, by more than 800 users who have taken more than 60,000 food images. For example, the Connecting Health and Technology (CHAT) study [11, 12] was a six-month randomized controlled trial (RCT) in 247 young adults (18-30 years). The study aimed to evaluate the effectiveness of tailored dietary feedback and weekly text messaging to improve dietary intake of fruit, vegetables and junk food over 6 months among a population-based sample of men and women (aged 18 to 30 years) [11, 12]. In this paper, we describe the current status of the TADA project and overview problems that still need to be addressed.

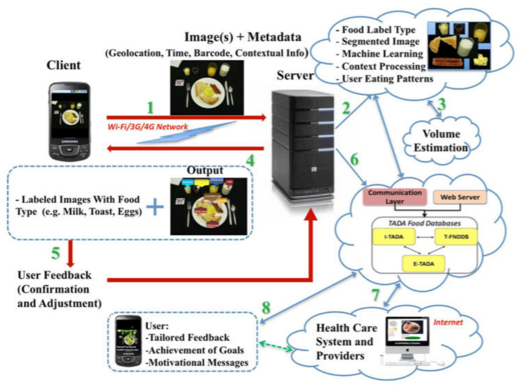


**Fig. 1**. The Architecture of the TADA System.

### 1.1. Other Work in Image-Based Dietary Assessment

In recent years, mobile-based "nutrition" systems and services have become increasingly popular. Some of these systems are capable of taking images of foods eaten at different eating occasions and using the images as part of a food diary to assist users in recording their diets. These systems include FoodLog [13], *Tuingle* [14], *FoodCam* [15], *DietCam* [16] and *Im2Calories* [17]. DietCam [16] uses images acquired from multiple views to analyze food. FoodLog [13] provides both a mobile application and cloud service that allows users to record daily dietary intake by acquiring images of food. In many of these systems, a user must first identify the names and quantities of food items and then the nutrient values are estimated, which places a large portion of the dietary assessment on the user (or on a human analyst). Tuingle [14] is able to automatically recognize food items, but it only works for images with a single food item and requires users to manually input portion size to complete energy assessment. *Im2Calories* [17], developed by Google, uses

This work was partially sponsored by the US National Institutes of Health under grant NIH/NCI 1U01CA130784-01 and NIH/NIDDK 1R01-DK073711-01A1, 2R56DK073711-04 and a Healthway Health Promotion Research Grant and from the Department of Health, Western Australia. 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 US National Institutes of Health. Address all correspondence to Edward J. Delp, ace@ecn.purdue.edu or see www.tadaproject.org.

deep learning techniques to recognize food but it has not been released to the public.

Besides the overall dietary collection systems mentioned above, there has been efforts in the area of food image analysis, in particular segmentation, food recognition and portion size estimation. The accurate estimation of energy and nutrients consumed using food images is essentially dependent on the correctly labelled food item and a sufficiently well-segmented region. Extensive research has been done to obtain reliable food segments and to learn a variety of food features and classifiers. Most of the work in food identification [18, 19, 20, 21] have analyzed multiple features and classifiers that are more effective to their applications. Some work [22, 23, 17] use either an end-to-end deep neural network or deep features with variations of SVM to obtain better results. In [22], the authors trained a CNN on the Food101 dataset [18] with 101 food categories and used domain adaption to improve the classification performance. [23] reports achieving the top 1 classification accuracy, 78.77% and 67.57% on the UEC-FOOD100/256 dataset.

Due to the complexity of food images (e.g. occlusion and cluttered background), food image segmentation is a difficult task. Some the food classification work mentioned above only deal with single food images, thus image segmentation is not necessary. For example, [20] focuses on restaurant foods and utilize GPS information to get restaurant menus. With the correct restaurant information, they can already map the dish name to the corresponding nutrient table. In [24], 10 types of food with containers were examined using a formable part model and a circle detector to constrain the food region of interest. Bettadapura et. al [21] used hierarchical segmentation in their implementation. A semantic segmentation method based on deep neural network was recently proposed in [17].

We define portion size estimation as the process of determining how much food (in $cm^3$ or grams) is present in the food image. Food volume estimation (or portion size estimation) is a challenging problem since the food preparation process and the way food is consumed can cause large variation in food shape and appearance. Many existing image-based work on food volume estimation require either modifying the mobile device such as 3D range finding [25], acquiring multiple images [26, 16], or video [27] which is not desirable for users trying to collect information about their diets and can contribute to poor compliance with these methods. Our work [10] has focused on the use of a single image for portion estimation since our studies have indicated that this reduces users' burden from having to acquire multiple images of their eating occasions.

## 2. TADA SYSTEM OVERVIEW AND CURRENT STATUS

### 2.1. Image Segmentation

Food image segmentation has been extensively addressed in our previous work resulting in a joint segmentation/classification multiple hypothesis technique [7]. We also investigated the use of local variation [28] and integrated it with food classifiers so that we can iteratively use the classification results to refine the segmented regions [29]. Local variation is a graph based segmentation method, in which two regions are segmented if the difference between the two regions 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 $\beta$ [28].

Since the image segmentation method is limited by a particular choice of input parameters, some food items may be under-segmented, while others may be over-segmented. We seek to overcome the segmentation problem by using classification feedback to refine the segmentation results. In our earlier work, Zhu [7] used a set of segmentation parameters for salient regions for segmentation refinement. Ye [29] proposed to use the initial segmentation result to replace the saliency region detection. After obtaining the initial segmented regions, segments which are smaller than $1/50$ of the original image are removed. Then each remaining segment is re-segmented and classified again. If the food classification confidence score is improved by re-segmentation, we accept the new segmentation; otherwise the original segmentation is kept as final segmentation. For each adjacent pair of segments, if a food category label in one segment equals to a label in the other segment, and the sum of the confidence score is greater than the highest individual score, we merge these two segments.

Although the segmentation refinement approach boosts the food segmentation accuracy, it is time-consuming due to multiple iterations of segmentation and classification. Also, general graph-based segmentation methods use low level features to measure similarity between two sets of pixels. For an image of a complicated scene, using low level features often result in noisy segments. To address the above disadvantages, we proposed an superpixel based normalized cut (SNcut) [30]. The idea is to obtain higher level features from superpixels and in the meantime reduce the size of the affinity matrix in Normalize Cut (Ncut) [31].

In [30], we use simple linear iterative clustering [32] to generate an initial segmentation mask. To construct a weighted graph connecting the superpixels, we used average RGB values and a customized local binary pattern [30] as the color and texture cues for the superpixels and $\chi^2$ test [30] in texture space and $L_2$ distance in color space between adjacent superpixels to determine the weights. We then map the similarity measurement to a probability estimate. Finally, the edge weight is obtained by combining the probability estimates from both features. To evaluate the proposed segmentation methods, we described a new metric in [30]. Based on our experiments, SNcut outperforms some of the widely used segmentation methods and it also produces competitive results for natural images based on other segmentation benchmarks [30].

### 2.2. Food Classification

Once the food items are segmented for a eating occasion image, color, texture and local region features are extracted for each segment. We have trained multiple classifiers and investigated the effective combination of features. Features are used for describing the characteristics of objects. An essential step in solving the food classification problem is to select suitable features to distinguish one food from another. We overview three types of features we have investigated extensively in the TADA system: color, texture and local region features, among which we regard color and texture as global features [7, 8].

We used two color descriptors, namely, Dominant Color Descriptor (DCD) and Scalable Color Descriptor (SCD) [8]. DCD is a vector of $D$ representative colors from the *CIE-Luv* color space using the generalized Lloyd algorithm for color clustering and their corresponding percentages [8]. SCD is determined by quantizing the colors in the *HSV* color space uniformly into 256 bins, which includes 16 levels in *H*, 4 levels in *S*, and 4 levels in *V* as suggested by the MPEG-7 standard [33].

We investigated two texture descriptors for food classification: Entropy-Based Categorization and Fractal Dimension Estimation (EFD) and Gabor-Based Image Decomposition and Fractal Dimension Estimation (GFD) [7]. EFD can be seen as an attempt to characterize the variation of roughness of homogeneous parts of the texture in terms of complexity [7]. GFD is based on fractal dimension estimation [7]. Instead of using entropy categorization, Gabor-based image decomposition is used. The fractal dimension descriptor is estimated for each filtered response and fused into one feature vector.

Local region features are described for points of interest and/or local regions. We used Scale Invariant Feature Transforms (SIFT) [34] and Multi-scale Dense SIFT (MDSIFT) in our previous work[7].

We use K-Nearest Neighbors (KNN) for the color and texture features and the Vocabulary Tree (VT) classifier for the local features [8]. Finally, we classify each segment using the weighted sum of confidence scores from KNN classifiers for each global features and from the vocabulary tree for local features. We tested the proposed method on the dataset of 1,453 eating occasion images of 42 commonly eaten food items and we achieved approximately 64% and 85% classification accuracy for Top 1 and Top 4 results [8].

### 2.3. The Use of Contextual Information

In our previous work [35, 9], we described the idea of using contextual information to refine the classification result, because often it is visually impossible to differentiate similar food items such as diet soda vs regular soda and a nonfat milk vs whole milk. By incorporating temporal and food co-occurrence patterns within a personalized learning model, we were able to deal with such issues effectively.

Contextual information, referring to any information that is not directly produced by the visual appearance of an object, has been discussed a great deal in image analysis and computer vision in the past few years [36, 37, 38]. In terms of food images, we investigated using temporal information and food co-occurrence patterns to improve food classification accuracy. Other contextual information we are investigating includes the location where the food has been consumed, data from the embedded pedometer and accelerometer, weather condition and information about the person's diet.

To gradually learn a person's daily diet, we proposed to use a recursive Bayesian model [35, 9]. The learning process is achieved by incorporating user feedback from our mobile interface, which will be discussed in Section 2.5. The food co-occurrence pattern is learned from a study in the wild [5], in which we provided common foods to the participants and the participants were also allowed to eat their own food according to their preference in terms of how and when they eat.

A food co-occurrence pattern describes the likelihood of food combinations. We use a graphical model to visualize the probabilistic structure. Since the number of food segments in an eating occasion is relatively small, we construct a weighted complete digraph between all segments. Experimental results showed the classification accuracy was improved by 11% on average after contextual refinement. More details are available in [35, 9].

### 2.4. Portion Size Estimation

Our ultimate goal is to determine from the food image the energy (kilocalories) and nutrients consumed. To do this we need to estimate size of the food items. This means we need to estimate the volume of each food item. Once we know this, we can use a food composition database such as the FNDDS [39, 40]. We have developed a single-view based food portion estimation technique [10]. Although single-view 3D scene reconstruction is in general an ill-posed problem, the use of geometric models such as the shape of a container can help to partially recover 3D parameters of the food items in the scene. The correct food classification label and segmentation mask in the image is alone insufficient for 3D reconstruction of a food item, hence the use of geometric models will allow for volume estimation where we can use the food label to index into a class of geometric models for single view volume estimation. The task then becomes finding the correct parameters for the selected geometric model. Based on the estimated 3D parameters of each food item and a reference object (a fiducial marker [10]) in the scene, the weight of each food can then be estimated using the estimated volume and density of the food item.

For example, the most commonly used models that have significant 3D structure either can be modeled as cylinders or can be approximated to be cylinders. For a cylinder, radius and height are sufficient to estimate the volume. Instead of reconstructing the cylinders, we designed an iterative point search technique to estimate the essential parameters which carry sufficient information for volume estimation. The iterative point search technique is based on projecting points from world coordinates to pixel coordinates, where the projection process is made possible using camera intrinsic and extrinsic parameters [10]. For foods served in non-rigid shapes or do not have significant 3D structures, we use the prism model. The prism model is an area-based model assuming the height is the same for the entire horizontal cross-section [10]. In an experiment using 19 food categories, we achieved 6% error rate for energy estimation assuming perfect segmentation mask and food classification [10].

### 2.5. The TADA Mobile Application

We have designed a mobile application for the iOS and Android mobile telephones that allow users to take images of their food and then submit these images to our system for analysis (see Figure 1). We refer to this mobile application as the mobile food record (mFR). The mFR has a simple, easy to use, GIU that has been extensively tested in our user studies [41].

The use of image analysis imposes requirements on the mFR and made the design of the mFR challenging. The current mFR has two main components: record mode and review mode [41]. The record mode is used for capturing images of eating occasions, while the review mode is used for reviewing the results of the food image classification and segmentation.

### 2.6. User Studies

The TADA system has been tested, validated and used globally by researchers, dietitians and nutritionists for various purposes. We have more than 14 user studies involving more than 800 users who acquired more than 60,000 food images under controlled and community-dwelling conditions [42, 43, 11, 12, 44]. Most of our studies showed that not only did we reduce the burden of collecting daily dietary food records on users but we also allowed health care professionals to have real-time access to the records.

In [42] we studied the ability of adolescents aged 11-18 years to identify foods in images of their meals 10-14h postprandial and estimate portion size. We showed that the automated processes performed better at estimating portion size than the adolescents. The TADA system examined the consumption of razor clam [44] with respect to domoic acid consumption which can be toxic for humans. A study described in [43] recruited 135 volunteers (78 adolescents, 57 adults) to use the mFR for one or two meals under controlled conditions in order to evaluate the set of skills among adolescents and adults to use mFR and to compare their preference regarding to mFR. The results show that most of the users are able to easily use the mFR, while the adults were more likely than adolescents to remember to capture images and include all foods and beverages in their images, but they were less efficient than adolescents to capture a satisfactory image. In the Connecting Health and Technology (CHAT) study [11, 12], 247 young adults aged 18-30 years participated in a 6-month study in a randomized controlled trial (RCT). The study aimed to evaluate the effectiveness of tailored feedback through the mFR to make dietary habit changes [11, 12]. Result shows that tailored dietary feedback have an important effect on reducing sugar-sweetened beverages and energy dense nutrient poor foods such as fast food, as well as reducing body weight in those who were overweight.

## 3. CONCLUSION AND FUTURE WORK

Food image analysis is an interesting and difficult problem and is attracting the interest of the larger computer vision and machine leaning community. With respect to the our TADA system, we intend to continue investigating both food portion size estimation and food image analysis. Estimating nutrient information or energy consumption from a single image is a very challenging research problem. In the future we would also like to incorporate contextual information for volume estimation. Deep learning has been gaining widespread attention and constitutional neural network has demonstrated impressive successes on many computer vision tasks. Compared to common objects, e.g. human, faces or buildings, food items in general lack structure. We are working on designing and implementing an effective food localization network and a localization to classification to improve food identification accuracy. In the future, we are also interested in analyzing the hierarchies of the common food categories and using crowdsourcing and user feedback to train our system more efficiently.

## 4. REFERENCES


[1] B. Six, T. Schap, F. Zhu, A. Mariappan, M. Bosch, E. Delp, D. Ebert, D. Kerr, and C. Boushey, "Evidence-based development of a mobile telephone food record," *Journal of American Dietetic Association*, vol. 110, pp. 74–79, January 2010.

[2] T. F. Aflague, C. J. Boushey, R. T. L. Guerrero, Z. Ahmad, D. A. Kerr, and E. J. Delp, "Feasibility and use of the mobile food record for capturing eating occasions among children ages 3–10 years in guam," *Nutrients*, vol. 7, no. 6, pp. 4403–4415, 2015.

[3] "Smartphone ownership and internet usage continues to climb in emerging economies," Pew Research Center. [Online]. Available: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

[4] 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, pp. S50–S57, February 2009.

[5] T. Schap, F. Zhu, E. Delp, and C. Boushey, "Merging dietary assessment with the adolescent lifestyle," *Journal of Human Nutrition and Dietetics*, vol. 27, no. s1, pp. 82–88, 2014.

[6] F. Zhu, M. Bosch, I. Woo, S. Kim, C. Boushey, D. Ebert, and E. 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] F. Zhu, M. Bosch, N. Khanna, C. Boushey, and E. Delp, "Multiple hypotheses image segmentation and classification with application to dietary assessment," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 377–388, January 2015.

[8] Y. He, C. Xu, N. Khanna, C. Boushey, and E. Delp, "Analysis of food images: Features and classification," *Proceedings of the IEEE International Conference on Image Processing*, pp. 2744–2748, October 2014, Paris, France.

[9] Y. Wang, Y. He, F. Zhu, C. Boushey, and E. Delp, "The use of temporal information in food image analysis," *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, ser. Lecture Notes in Computer Science, V. Murino, E. Puppo, D. Sona, M. Cristani, and C. Sansone, Eds. Springer International, 2015, vol. 9281, pp. 317–325.

[10] S. Fang, C. Liu, F. Zhu, E. Delp, and C. Boushey, "Single-view food portion estimation based on geometric models," *Proceedings of the IEEE International Symposium on Multimedia*, pp. 385–390, December 2015, Miami, FL.

[11] D. A. Kerr, C. Pollard, P. A. Howat, E. J. Delp, M. Pickering, K. R. Kerr, S. S. Dhaliwal, I. S. Pratt, J. Wright, and C. J. Boushey, "Connecting health and technology (chat): Protocol of a randomized controlled trial to improve nutrition behaviours using mobile devices and tailored text messaging in young adults," *BMC Public Health*, vol. 12, no. 447, pp. 1–10, June 2012.

[12] D. A. Kerr, A. J. Harray, C. M. Pollard, S. S. Dhaliwal, E. J. Delp, P. A. Howat, M. R. Pickering, Z. Ahmad, X. Meng, I. S. Pratt, J. L. Wright, K. R. Kerr, and C. J. Boushey, "The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support in young adults," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 13, no. 1, p. 52, April 2016.

[13] K. Kitamura, T. Yamasaki, and K. Aizawa, "Foodlog: Capture, analysis and retrieval of personal food images via web," pp. 23–30, November 2009.

[14] "Tuingle." [Online]. Available: http://tuingle.com/

[15] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning," *Proceedings of the IEEE International Conference on Image Processing*, pp. 285–288, October 2009, cairo, Egypt.

[16] F. Kong, H. He, H. A. Raynor, and J. Tan, "Dietcam: Multiview regular shape food recognition with a camera phone," *Pervasive and Mobile Computing*, vol. 19, pp. 108–121, 2015.

[17] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. P. Murphy, "Im2calories: Towards an automated mobile vision food diary," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1233–1241, December 2015, Santiago, Chile.

[18] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 – mining discriminative components with random forests," *European Conference on Computer Vision*, vol. 8694, pp. 446–461, September 2014.

[19] Y. Kawano and K. Yanai, "Foodcam-256: a large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights,"

*Proceedings of the ACM international conference on Multimedia*, pp. 761–762, November 2014.

[20] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, "Menu-match: Restaurant-specific food logging from images," *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 844–851, January 2015, Waikoloa Beach, HI.

[21] V. Bettadapura, E. Thomaz, A. Parnami, G. Abowd, and I. Essa, "Leveraging context to support automated food recognition in restaurants," *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 580–587, January 2015, Waikoloa Beach, HI.

[22] S. Ao and C. X. Ling, "Adapting new categories for food recognition with deep representation," *Proceedings of the IEEE International Conference on Data Mining Workshop*, pp. 1196–1203, 2015.

[23] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," *Proceedings of the IEEE International Conference on Multimedia & Expo Workshops*, pp. 1–6, 2015.

[24] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 25–30, July 2012, Melbourne, Australia.

[25] J. Shang, M. Duong, E. Pepin, X. Zhang, K. Sandara-Rajan, A. Mamishev, and A. Kristal, "A mobile structured light system for food volume estimation," *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 100–101, November 2011, Barcelona, Spain.

[26] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pp. 1–8, December 2009, Snowbird, UT.

[27] M. Sun, J. Fernstrom, W. Jia, S. Hackworth, N. Yao, Y. Li, C. Li, M. Fernstrom, and R. Sclabassi, "A wearable electronic system for objective dietary assessment," *Journal of the American Dietetic Association*, p. 110(1): 45, January 2010.

[28] P. Felzenszwalb and D. Huttenlocher, "Image segmentation using local variation," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 98–104, June 1998, Santa Barbara, CA.

[29] Y. He, C. Xu, N. Khanna, C. Boushey, and E. Delp, "Food image analysis: Segmentation, identification and weight estimation," *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1–10, July 2013, San Jose, CA.

[30] Y. Wang, C. Liu, F. Zhu, C. J. Boushey, and E. J. Delp, "Efficient superpixel based segmentation for food image analysis," *Proceedings of the IEEE International Conference on Image Processing*, September 2016, to appear.

[31] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, August 2000.

[32] 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*, vol. 34, no. 11, pp. 2274–2282, November 2012.

[33] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703 –715, June 2001.

[34] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 2, no. 60, pp. 91–110, January 2004.

[35] Y. He, C. Xu, N. Khanna, C. Boushey, and E. Delp, "Context based food image analysis," *Proceedings of IEEE International Conference on Image Processing*, pp. 2748–2752, September 2013, Melbourne, Australia.

[36] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," *Proceedings of IEEE International Conference on Computer vision*, pp. 1–8, October 2007, Rio de Janeiro, Brazil.

[37] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Computer Vision and Image Understanding*, vol. 114, pp. 712–722, February 2010.

[38] B. McFee, C. Galleguillos, and G. Lanckriet, "Contextual object localization with multiple kernel nearest-neighbor," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 570–585, February 2011.

[39] "USDA food and nutrient database for dietary studies, 1.0." Beltsville, MD: Agricultural Research Service, Food Surveys Research Group, 2004.

[40] S. Kelkar, S. Stella, C. Boushey, and M. Okos, "Developing novel 3D measurement techniques and prediction method for food density determination," *Procedia Food Science*, vol. 1, pp. 483 – 491, May 2011.

[41] Z. Ahmad, M. Bosch, N. Khanna, D. A. Kerr, C. J. Boushey, F. Zhu, and E. J. Delp, "A mobile food record for integrated dietary assessment," *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pp. 53–62, October 2016, amsterdam, Netherlands.

[42] T. E. Schap, B. L. Six, E. J. Delp, D. S. Ebert, D. A. Kerr, and C. J. Boushey, "Adolescents in the united states can identify familiar foods at the time of consumption and when prompted with an image 14 h postprandial, but poorly estimate portions," *Public Health Nutrition*, vol. 1, no. 1, pp. 1–8, July 2011.

[43] B. L. Daugherty, T. E. Schap, R. Ettienne-Gittens, F. Zhu, M. Bosch, E. J. Delp, D. S. Ebert, D. A. Kerr, and C. J. Boushey, "Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents," *Journal of Medical Internet Research*, vol. 14, no. 2, p. e58, April 2012.

[44] C. J. Boushey, E. J. Delp, Z. Ahmad, Y. Wang, S. M. Roberts, and L. M. Grattan, "Dietary assessment of domoic acid exposure: what can be learned from traditional methods and new applications for a technology assisted device," *Harmful Algae*, vol. 57, pp. 51–55, July 2016.