Quantification of Food Intake Using Food Image Analysis

Corby K. Martin, Sertan Kaya, Bahadir K. Gunturk

Abstract—**Measuring free-living peoples' food intake represents methodological and technical challenges. The Remote Food Photography Method (RFPM) involves participants capturing pictures of their food selection and plate waste and sending these pictures to the research center via a wireless network, where they are analyzed by Registered Dietitians to estimate food intake. Initial tests indicate that the RFPM is reliable and valid, though the efficiency of the method is limited due to the reliance on human raters to estimate food intake. Herein, we describe the development of a semi-automated computer imaging application to estimate food intake based on pictures captured by participants.**

Index Terms—**Food intake; Energy intake; Computer imaging; Digital photography; Remote Food Photography Method (RFPM)**

I. INTRODUCTION

OBTAINING real-time and accurate estimates of food intake while people reside in their natural environment is technically and methodologically challenging. Currently, doubly labeled water is the "gold standard" for measuring food intake in free-living conditions. Doubly labeled water provides a measure of total daily energy expenditure (TEE), and during energy balance (weight maintenance), energy expenditure equals energy (food) intake [1, 2]. If weight is not stable or energy balance is not present during the period of measurement, TEE data are adjusted for change in energy stores [3]. Nevertheless, it is difficult to obtain an accurate estimate of an individual's food intake using doubly labeled water if the person experiences a large energy deficit during the period of measurement, even if change in energy stores is considered [3]. Consequently, doubly labeled water has limitations when people are dieting, or the times at which clinicians frequently require accurate estimates of food intake. Additional limitations of doubly labeled water include its cost, inability to provide data on the types of foods consumed, and the lack of real-time data.

C. Martin is with the Pennington Biomedical Research Center, Louisiana State University-System, Baton Rouge, LA 70808 USA (phone: 225-763-2585; fax: 225-763-3045; e-mail: martinck@pbrc.edu). S. Kaya and B. K. Gunturk are with the Dept. of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803 USA (phone: 225-578-5241; fax: 225-578-5200; email: skaya1@lsu.edu and bahadir@lsu.edu).

Most alternative methods to measure food intake rely on participant self-report, including food records, 24-hour recall, and food frequency questionnaires. These methods rely on the participant to recall or record the foods that they eat and estimate or measure the amount (portion) of food eaten. Although these methods are frequently utilized to estimate food intake in research and clinical settings, they underestimate food intake by 37% or more [4-6]. In addition, people who are overweight or obese underreport food intake to a greater degree than lean people when using these methods [6]. The largest source of error in estimating food intake from self-report is attributable to participants' poor estimation of portion size [7]. Hence, methods that do not rely on the participant to estimate portion size are needed.

The digital photography of foods method [8, 9] was developed to unobtrusively measure food intake in cafeteria settings and it does not rely on the participant to estimate portion size. The digital photography of foods method involves using a digital video camera to capture a photograph of a participant's food selection before they eat, and plate waste after they finish eating. While in the cafeteria or dining location, photographs are also captured of carefully measured standard portions of the foods served on the day of data collection. At a later date in the laboratory, these photographs are analyzed by registered dietitians (RDs) who estimate the amount (portion) of food selection and plate waste by comparing these photographs to the standard portion photographs. These portion size estimates are entered into a custom built computer application that automatically calculates the grams, kilocalories (kcal), and macro- and micro-nutrients of food selection, plate waste, and food intake based on a United States Department of Agriculture (USDA) database [10]. The digital photography of foods method has been found to be highly reliable and accurate (valid) and overestimates food intake by less than 6 grams on average [9].

The digital photography of foods method represents a novel and valid method for quantifying food intake in cafeteria settings, but it is not appropriate for free-living conditions. The Remote Food Photography Method (RFPM) [11], however, was developed specifically to measure food intake in free-living conditions and it builds upon the digital photography methodology. When using the RFPM, participants are trained to use a camera-equipped cell phone with wireless data transfer capabilities to take pictures of their food selection and plate waste and to send these pictures to the researchers over the wireless network. To reduce the frequency of participants forgetting to take photographs of their foods, they receive and respond to automated prompts reminding them to take photographs and

to send the photographs to the researchers. These prompts are consistent with ecological momentary assessment (EMA) [12] methods and consist of emails and text messages. The images are received by the researchers in near real-time and can be analyzed quickly to estimate food intake. When analyzing the images, the RDs rely on methods similar to the digital photography of foods method, but the participants are not required to take a photograph of a standard portion of every food that they eat, as this would be unfeasible. Alternatively, an Archive of over 2,100 standard portion photographs was created. This allows the RDs to match foods that participants eat to a standard portion photograph that already exists. Initial tests supported the reliability and validity of the RFPM [11]. The RFPM underestimated food intake by ~6% and, importantly, the error associated with the method was consistent over different levels of body weight and age.

The purpose of the research reported herein was to develop an automated food intake evaluation system (AFIES) to: 1) identify the foods in pictures of food selection and plate waste, and 2) automatically calculate the amount of food depicted in photographs of food selection and plate waste. Hence, the AFIES would replace the RDs who manually estimate the amount of food selection and plate waste, making the RFPM much more efficient and cost effective, and possibly more accurate.

II. AUTOMATED EVALUATION OF FOOD INTAKE

The automated food intake evaluation system consists of reference card detection, food region segmentation and classification, and food amount estimation modules.

A. Reference Card Detection

In order to estimate food portions accurately in free-living conditions, we need a reference in the pictures to account for the viewpoint and distance of the camera. For this purpose, the subjects are asked to place a reference card next to their food before taking a picture. The reference card is a standard 85.60x53.98mm [ISO/IEC 7813] ID card with a specific pattern printed on top. The pattern consists of two concentric rectangle (bull's-eye) patterns and a surrounding rectangle, as seen in Fig.1 and Fig.2. The bull's-eye patterns are used to locate the reference card within the picture; and the surrounding rectangle is utilized for determining the four corners of the card.

The first step in reference card detection is to binarize the input image. Since global thresholding is likely to fail in capturing the local structures (therefore, the bull's-eye pattern), we employ an adaptive thresholding method [13]. We take the difference between the luminance channel of an image and its filtered version (which is obtained using an 11x11 averaging filter), and threshold the difference image to obtain the binary image. The pattern detector starts with the first row of the binary image, runs along the rows, first from left to right and then from right to left, and returns a high value at center locations of alternating color patterns with mirror symmetry. Such bull's-eye detection is also used in localization of 2D barcodes, such as the MaxiCode [14].
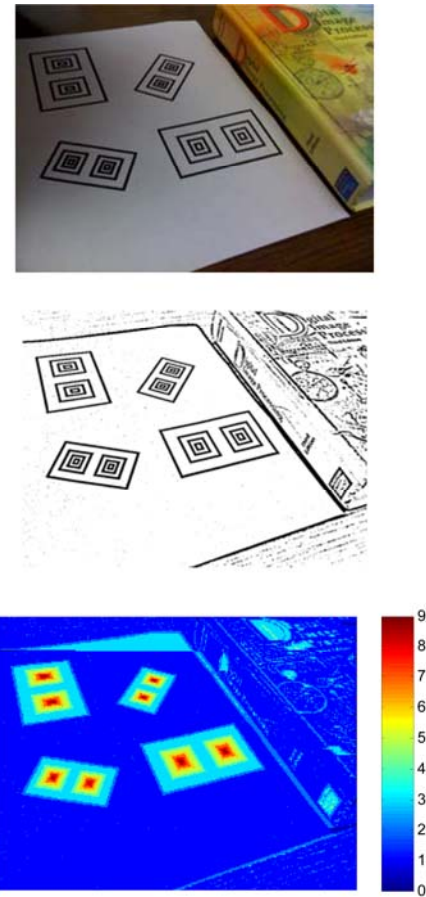

Fig.1. Reference card detection. Top: Input image; Middle: Result of adaptive thresholding; Bottom: Response of the pattern detector.

Fig.1 illustrates that the adaptive thresholding method successfully extracts the local texture even if there is non-uniform illumination, and the reference card detector works well regardless of the orientation or the perspective of the card.

Once the patterns are located, we do a morphological region fill operation on the binary image to determine the exact location of the entire reference card. The seed points of the region fill operation are chosen as the points along the line that connects the centroids of the two peak regions (bull's eye centers) inside the card; this guarantees filling of the entire card region. This is followed by the Harris corner detector [15] to locate the four corners of the card, which can later be used for perspective correction of the food area estimates. An example is shown in Fig.2.

B. Food Region Segmentation and Classification

The next step is the segmentation and classification of the food in the picture. This requires extraction of the features for each food type in a training process. During training, a user manually selects the food region for each food type; the features associated with that food type are extracted and saved. (Fig.3 shows a screenshot of the manual region selection process.) The user is also asked to enter the gram amount (or volume) of the food type; this information is used to establish the association between the food region

Fig.2. Corner detection and perspective correction. In raster scan order: Binary image after adaptive thresholding; Detected reference card; Corners of the card are marked on the image; Perspective correction of the card.
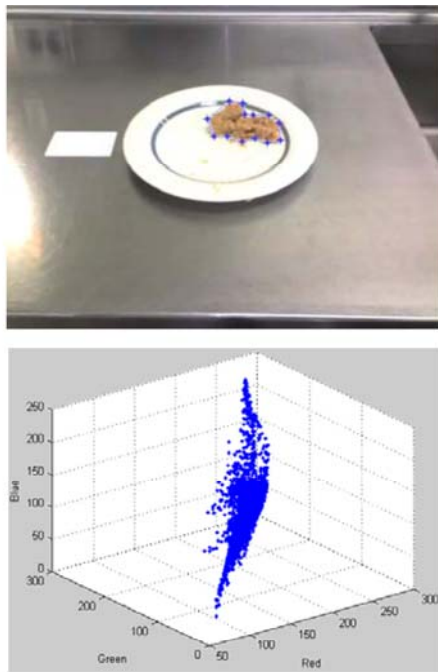

Fig.3. Manual selection of the food region during the training stage. Top: The region of interest is outlined by the blue points that are clicked by the user. Bottom: Distribution of the color samples in RGB space.

area and the gram amount for each food type. After completing the process for all food types, a classifier is trained given the features. In our prototype system, we use the color RGB data (red, green, and blue values) as the feature vector for the classifier. In future versions, we will expand the feature space with additional features, such as the Gabor texture features [16], and apply a dimensionality reduction technique, e.g. the Principal Component Analysis (PCA), to obtain the principal components. The distribution of the RGB data in a selected food region is modeled as a Gaussian distribution; the parameters of the Gaussian model, namely, mean and covariance matrix are then computed for

that particular food type. (We should also note here that the white regions of the reference card are used to do white balancing, which improves the performance.) During the testing stage, the distance between each pixel to a food class is calculated using the Mahalanobis distance [17], which accounts for the distribution of the feature values. A typical distance image is shown in Fig.4. The distance calculation process is repeated for all food classes. Pixels that are close to a class (i.e., pixels with Mahalanobis distance less than a pre-determined threshold) are assigned to that class. After repeating for all classes, we have the food regions for each class. In our current system, we restrict one food type per image; therefore, we pick the food class with the largest area. We then do a refinement step, where region growing and morphological denoising (opening and closing) operations are performed to obtain the final food region. Fig.4 shows a successful segmentation/classification result. For the future versions of our system, we will investigate the use of different classification techniques. Support Vector Machines [18], multilayer neural networks [19], and classifier fusion are among the techniques we are considering. These techniques can be incorporated into the classification module of our system without any special treatment.
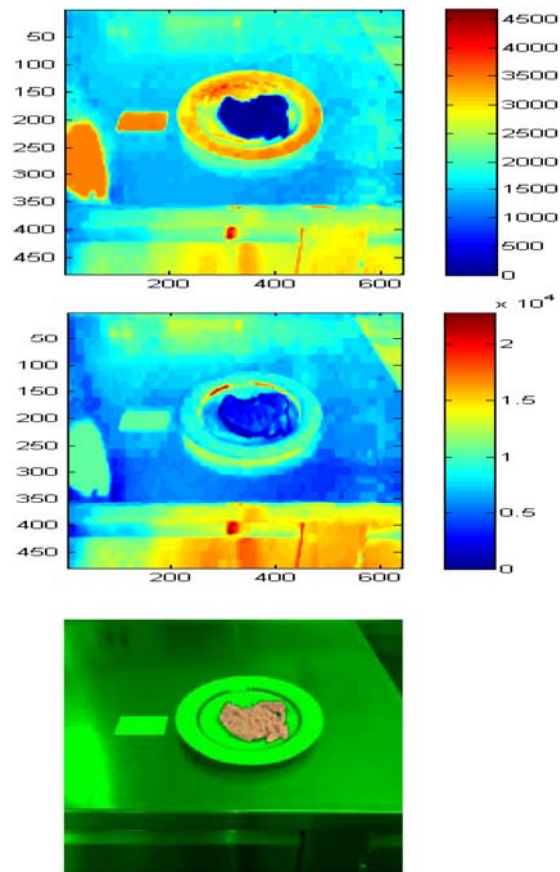

Fig.4. Automatic region segmentation and classification. Top: Mahalanobis distance of pixels to the food class shown in Fig.3. Middle: Mahalanobis distance of the pixels to another food class. Bottom: Final segmented and classified region.

C. Food Amount Estimation

As described earlier, the real area of the food region is determined using the reference card. Based on the association between the food region area and gram amount (equivalently, volume), the amount of food in the picture is estimated. The formula between the area and the volume depends on the food type and shape of the plate used. For some food, the volume is roughly proportional to the area. On the other hand, for some other foods, such as soup, the shape of the bowl needs to be known to establish the formula between the area and the volume. (See Fig.5 for an illustration.) In our current system, we assume linear proportionality between the food area and the volume. As a result, we estimate the area (therefore, the volume and gram amount) of the food in before and after pictures, and estimate the amount of food intake. We will add a feature to our system to associate the area-volume formula with the food type so that the volumes are estimated more accurately for food that is put in bowls; the system will assume a standard bowl shape to establish the area-volume formula.
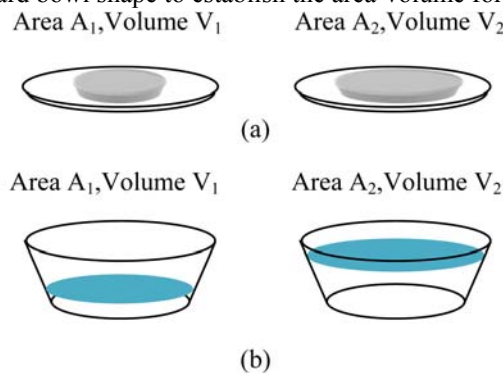

Fig.5. Area-Volume relation. (a) Volume is linearly proportional to the area; that is, $V_2 = A_2 (V_1 / A_1)$. (b) The volume depends on the shape (namely, top and bottom areas and the height) of the bowl, which is modeled as a cut cone.

D. Manual Review

The automated system estimates the food type and gram amount in each image and saves the information. This long process is performed offline. A dietitian then reviews the results and if necessary changes estimated gram amounts manually. This control mechanism guarantees the accuracy of the data to be used in food intake analysis.

III. CONCLUSIONS AND FUTURE WORK

This paper presents a framework to automatically estimate the food intake in free living conditions. We use a reference card system to estimate the true area of the food portions. The system relies on accurate detection of the corners of the reference. In the event of low-resolution images, the corners may not be estimated accurately; in such a case, we may do a template matching to the reference card or fit lines to the edges of rectangles in the card using Hough transform to eventually have a more accurate estimate of the position of the card. Currently, we are using color features in our system; this is obviously not sufficient as different foods may have similar color features. We will add texture features to improve the performance of our system. Another addition to our system will be an advanced classifier instead of the Mahalanobis distance classifier, which is currently employed. We are also planning to include image enhancement modules, such image denoising, compression artifact reduction, and contrast enhancement subsystems, to handle low-quality images. Another future work is to investigate the use of multiple images to construct 3D structures, and to have more accurate estimate of the gram amounts.

REFERENCES

[1] M. B. Livingstone and A. E. Black, "Markers of the validity of reported energy intake," *J Nutr,* vol. 133 Suppl 3, pp. 895S-920S, Mar 2003.

[2] D. A. Schoeller, "How accurate is self-reported dietary energy intake?," *Nutr Rev,* vol. 48, pp. 373-9, Oct 1990.

[3] L. de Jonge, J. P. DeLany, T. Nguyen, J. Howard, E. C. Hadley, L. M. Redman, and E. Ravussin, "Validation study of energy expenditure and intake during calorie restriction using doubly labeled water and changes in body composition," *Am J Clin Nutr,* vol. 85, pp. 73-9, Jan 2007.

[4] A. H. Goris, M. S. Westerterp-Plantenga, and K. R. Westerterp, "Undereating and underrecording of habitual food intake in obese men: selective underreporting of fat intake," *Am J Clin Nutr,* vol. 71, pp. 130-4, Jan 2000.

[5] L. G. Bandini, D. A. Schoeller, H. N. Cyr, and W. H. Dietz, "Validity of reported energy intake in obese and nonobese adolescents," *Am J Clin Nutr,* vol. 52, pp. 421-5, Sep 1990.

[6] D. A. Schoeller, L. G. Bandini, and W. H. Dietz, "Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method," *Can J Physiol Pharmacol,* vol. 68, pp. 941-9, Jul 1990.

[7] J. Beasley, W. T. Riley, and J. Jean-Mary, "Accuracy of a PDA-based dietary assessment program," *Nutrition,* vol. 21, pp. 672-7, Jun 2005.

[8] D. A. Williamson, H. R. Allen, P. D. Martin, A. Alfonso, B. Gerald, and A. Hunt, "Digital photography: a new method for estimating food intake in cafeteria settings," *Eat Weight Disord,* vol. 9, pp. 24-8, Mar 2004.

[9] 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," *J Am Diet Assoc,* vol. 103, pp. 1139-45, Sep 2003.

[10] USDA, "United States Department of Agriculture, Agricultural Research Service, Continuing Survey of Food Intakes by Individuals, 1994-1996, 1998," 2000.

[11] C. K. Martin, H. Han, S. M. Coulon, H. R. Allen, C. M. Champagne, and S. D. Anton, "A novel method to remotely measure food intake of free-living people in real-time: The Remote Food Photography Method (RFPM)," *Br J Nutr,* vol. 101, pp. 446-456, 2009.

[12] A. A. Stone, & Shiffman, S., "Ecological Momentary Assessment (EMA) in behavioral medicine," *Ann Behav Med,* vol. 16, 1994.

[13] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*: Prentice-Hall, 2008.

[14] D. G. Chandler, E. P. Batterman, and G. Shah, "Hexagonal, Information Encoding Article, Process and System," US Patent No. 4874936.

[15] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. of the 4th Alvey Vision Conf.,* pp. 147-151, 1988.

[16] B. S. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, pp. 837-842, 1996.

[17] P. C. Mahalanobis, "On the Generalised Distance in Statistics," Proc. of the National Institute of Sciences of India, vol. 2, pp. 49-55, 1936.

[18] B. Scholkopf, C. J. C. Burges, and A. J. Smola, "Advances in Kernel Methods: Support Vector Learning," Cambridge, MA: MIT Press, 1999.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer, 2006.