Technology-Assisted Dietary Assessment

Fengqing Zhu, Anand Mariappan, Carol J. Boushey, Deb Kerr, Kyle D. Lutes, David S. Ebert, Edward J. Delp

School of Electrical and Computer Engineering
Department of Foods and Nutrition
Department of Computer and Information Technology
Purdue University, West Lafayette, Indiana USA
Curtin Institute of Technology
Perth, Western Australia

ABSTRACT
Dietary intake provides valuable insights for mounting intervention programs for prevention of disease. With growing concern for adolescent obesity, the need to accurately measure diet becomes imperative. Assessment among adolescents is problematic as this group has irregular eating patterns and have less enthusiasm for recording food intake. Preliminary studies among adolescents suggest that innovative use of technology may improve the accuracy of diet information from young people. In this paper, we propose a novel food record method using a mobile device that will provide an accurate account of daily food and nutrient intake among adolescents. Our approach includes the use of image analysis tools for identification and quantification of food consumption. Images obtained before and after food is consumed can be used to estimate the diet of an individual. In this paper we describe our initial results and indicate the potential of the proposed system.

Keywords: dietary assessment, diet record method, mobile telephone, food volume estimation image analysis, classification, pattern recognition, image texture, feature extraction

1. INTRODUCTION
The increasing prevalence of obesity among the youth is of great concern [1] and has been linked to an increase in type 2 diabetes mellitus [2]. Accurate methods and tools to assess food and nutrient intake are essential in monitoring the nutritional status of this age group for epidemiological and clinical research on the association between diet and health. The collection of food intake and dietary information provides some of the most valuable insights into the occurrence of disease and subsequent approaches for mounting intervention programs for prevention. The assessment of food intake in adolescents has been evaluated by a food record (FR), the 24-hour dietary recall (24HR), and a food frequency questionnaire (FFQ) with external validation by doubly-labeled water (DLW) and urinary nitrogen [3–7]. Currently, there are too few validation studies in children to justify one particular method over another for any given study design.

Assessment of diet among adolescents is problematic. Early adolescents, ages 11 to 14 years, in particular, are in that period of time when the novelty and curiosity of assisting in or self-reporting of food intakes starts to wane and the assistance from parents is seen as an intrusion [3]. Dietary assessment methods need to continue to evolve to meet these challenges. There is recognition that further improvements will enhance the consistency and strength of the association of diet with disease risk, especially in light of the current obesity epidemic among this group.

Preliminary studies among adolescents suggest that innovative use of technology may improve the accuracy of diet information from young people. Our goal is to develop, implement, and evaluate a mobile device (i.e., a PDA or mobile telephone) food record that will translate to an accurate account of daily food and nutrient intake among adolescents. Simply stated our goal is to use a mobile network connected device that has a camera to take

This work was sponsored granting from the National Institutes of Health under grants NIDDK 1R01DK073711-01A1 and NCI 1U01CA130784-01. Address all correspondence to E. J. Delp, ace@ecn.purdue.edu or see www.tadaproject.org.
images of food before it is consumed and estimate food intake using image analysis methods. Mobile computing devices provide a unique vehicle for collecting dietary information that reduces burden on record keepers. Images of food can also be marked with a variety of input methods that link the item for image processing and analysis to estimate the amount of food. Images acquired before and after foods are eaten can estimate the amount of food consumed.

In this paper, we describe some preliminary results from the development of such a system.

2. REVIEW OF CURRENT DIETARY ASSESSMENT METHODS

A review of some of the most popular dietary assessment methods are provided in this section. Our objective here is to analyze the advantages and major drawbacks of these methods. This will demonstrate the significance of our mobile system which can be used for population and clinical based studies to improve the understanding of dietary exposures among adolescents.

2.1. 24-Hour Dietary Recall

The 24-hour dietary recall (24HR) consists of a listing of foods and beverages consumed the previous day or the 24 hours prior to the recall interview. Foods and amounts are recalled from memory with the aid of an interviewer who has been trained in methods for soliciting dietary information. A brief activity history may be incorporated into the interview to facilitate probing for foods and beverages consumed. The Food Surveys Research Group (FSRG) of the United States Department of Agriculture (USDA) has devoted considerable effort to improving the accuracy of this method.

The major drawback of the 24HR is the issue of underreporting of the food consumed [8]. Factors such as obesity, gender, social desirability, restrained eating and hunger, education, literacy, perceived health status, age, and race/ethnicity have been shown to be related to underreporting [9–12]. Harnack, et al. [13] found significant underreporting of large food portions when food models showing recommended serving sizes were used as visual aids for respondents. Given that larger food portions have been observed as occurring over the past 20 to 30 years [14, 15] this may be a contributor to underreporting and methods to capture accurate portion sizes are needed. In addition, youth, in particular, are limited in their abilities to estimate portion sizes accurately [3].

The most common method of evaluating the accuracy of the 24HR with children is through observation of school lunch [16] and/or school breakfast [17] and comparing foods recalled with foods either observed as eaten or foods actually weighed. These recalls have demonstrated both under-reporting and over-reporting, and incorrect identification of foods.

2.2. The Food Record

The 24HR is useful in population based studies; the preferred dietary assessment method for clinical studies is the food record. Depending on the primary nutrient or nutrients or foods of interest, the minimum number of food records needed is rarely less than two days. Training the subjects, telephoning with reminders for recording, reviewing the records for discrepancies, and entering the dietary information into a nutrient database can take a large amount of time and requires trained individuals [18].

The food record is especially vulnerable to underreporting due to the complexity of recording food [19, 20]. A study among 10-12 year old children found significant underreporting of total energy intake (TEI) when the intake was compared against an external marker, doubly-labeled water (DLW) [21]. As adolescents snack frequently, have unstructured eating patterns, and consume greater amounts of food away from the home, their burden of recording will be much greater compared to adults. It has been suggested that these factors, along with a combination of forgetfulness and irritation and boredom caused by having to record intake frequently may be contributing to the underreporting in this age group [22]. Dietary assessment methods perceived as less burdensome and time-consuming may improve compliance [22].
2.3. Portion Size Estimation

Portion size estimation may be one contributor to underreporting. In [23] it was found that 45 minutes of training in portion-size estimation among 9-10 year olds significantly improved estimates for solid foods which were measured by dimensions or cups, and liquids estimated by cups. Amorphous foods were estimated least accurately even after training and some foods still exhibited an error rate of over 100%. Thus, training can improve portion size estimation, however more than one session may be needed.

2.4. Evaluation of Dietary Assessment Methods

The number of days needed to estimate a particular nutrient depends on the variability of the nutrient being assessed and the degree of accuracy desired for the research question [24–27]. Most nutrients require more than four days for a reliable estimate [25, 27]. However, most individuals weary of keeping records beyond four days which may decrease the quality of the records [19].

Another challenge in evaluating dietary assessment methods is comparing the results of the dietary assessment method to some measure of “truth.” This is best achieved by identifying a biomarker of a nutrient or dietary factor [20, 28]. The underlying assumption of a biomarker is that it responds to intake in a dose-dependent relationship [26]. The two methods that have widest consensus as valid biomarkers are DLW for energy [20, 29] and 24-hour urinary nitrogen for protein intake [6, 30, 31]. A biomarker does not rely on a self-report of food intake, thus theoretically the measurement errors of the biomarker are not likely to be correlated with those of the dietary assessment method. Other biomarkers collected from urine samples include potassium and sodium [30]. Plasma or serum biomarkers that have been explored are levels of ascorbic acid for vitamin C intake [30, 32], β-carotene for fruits and vegetables or antioxidants [32–34]. These latter markers are widely influenced by factors such as smoking status and supplement use, thus their interpretation to absolute intake is limited.

3. SYSTEM OVERVIEW

Simply stated our goal is to use a network connected mobile device that has a build-in camera to take images of food before and after it is consumed and estimate food intake using image analysis methods. Children and adolescents are the most eager in terms of adopting new technology. Mobile devices, such as PDAs and mobile telephones with cameras, are general purpose computing devices that have a great deal of computational power that can be exploited for solutions to this problem. PDAs are ideal as a field data collection device for diet assessment [35, 36]. However, there have been problems when deploying these types of devices if one does not understand the user and the environment in which the device will be deployed.

Researchers asked overweight and obese adults in a 24-week behavioral weight control program to self-report food intake using a PDA [37]. At the end of the study, almost half of the subjects were found to under-report their energy intakes suggesting that the PDA did not improve accuracy in these users. The PDA used in this study used old technology that lacked “user friendly options,” e.g., it lacked a back-lit display making the screen difficult to see. Therefore, the authors of that study concluded that the technology was a barrier.

We believe that a properly designed system will make mobile devices attractive to users and can be used to measure food intake. Most mobile devices include digital cameras which make taking pictures and labeling the content of the pictures less burdensome than writing on paper. The use of a mobile device that works the way young people interact with portable devices may address many of the issues outlined as barriers to recording food intake among adolescents. Young users treat their mobile device as an extension of their personality and this needs to be considered in the design of our system.

Mobile devices have the potential to create an entire new platform for applications and services that could be used for dietary assessments. For example, some individuals may forget to record their food when eaten, in which case the record can become a cross between a recall and a record. With paper and pencil recording there is no way a researcher can check that foods were recorded at the time of the meal or that all meals were recorded at the end of the day. With a mobile device, every entry records a time stamp, thus allowing researchers to more accurately determine if data entry occurred at typical meal times (record), long after the meal, or all at once at the end of the day (an unassisted recall). The use of an image provides another dimension of verifying food intake.
Figure 1. Examples of input techniques for use with mobile device food record (a) Using the tree method to select a food item from the database, (b) Using the search method to select a food item from the database, (c) Using the tree method to mark digital pictures with a food item in the database and (d) Using the search method to mark digital pictures with a food item in the database.

Figure 2. Additional input devices and labels for use with mobile device food record (a) Using stylus to hand write notes to label food items in a digital picture of a meal, (b) Using the onscreen tree method to label food items in a digital picture of a meal, (c) Using an onscreen keyboard to search for a food item from the database and (d) A labeled meal and not shown are the secondary prompts for dressing on salad and spread on bread.

Our goal is to use a mobile device with a build-in camera, integrate image analysis and visualization tools with a nutrient database, to allow an adolescent user to discretely “record” foods eaten. The user will capture an image of his/her meal or snack both before and after eating. We will develop automatic image analysis methods to determine the food consumed since this reduces the burden of many aspects of recording for the users and reduces analysis for the researchers. However, in the event that a picture of a food cannot be obtained, the system needs an alternative way of determining the food consumed. The user could then identify the foods consumed by writing on the screen, or by tapping the screen, and using various data entry menus and forms. Entry methods would include selecting foods from a tree list or searching for a food in the database using words or portions of words. Examples of these input techniques are shown in Figure 1 and Figure 2.

In addition, the device can prompt users with real-time reminders sent using the network capabilities of the mobile device. Reports can be sent to a central system to allow regular monitoring. The use of time-stamped food entries and images can aid the research dietitian or the clinical dietitian in reviewing the food record with or without the adolescent to identify foods.

3.1. Image Analysis and Visualization

We are developing methods to automatically estimate the food consumed at a meal from an image acquired from a mobile device. An example of such an image is shown in Figure 3. The goal is to identify each food item and estimate the volume of each food item consumed. From this information, the energy and nutrients consumed can be determined. This is not an easy problem in that some foods may not be identifiable from an image. For
example, the type of milk in a cup (e.g., low fat or skim milk) may not be determined from the image alone. This will require other types of “side information” be available to the system either through how the food was packaged (e.g., an image of the milk carton) or through inputs (manual or audio) from the user.

This requires that we have a calibrated, possibly 3D, imaging systems. A block diagram of our image analysis system is shown in Figure 4. Our plan for addressing the various tasks are described below.

![Figure 3. A Typical Image of a Meal.](image1)

![Figure 4. Image Analysis System.](image2)

**Image Calibration and Acquisition**

Since we are interested in knowing how much food was consumed, we need to have a 3D calibrated system. This could be accomplished by having the user take the image with a known fiducial object, e.g., a pen or PDA.
stylus, placed next to the food so one could use this to “calibrate sizes” in the images. We might also use the known dimensions of a plate or cup in a scene. Other a priori information in the scene such as the the pattern on the tablecloth (see Figure 3) could also be used.

For 3D or volume estimation we are also exploring the use of multiple images of the scene taken at different orientations. This will also require that calibration information be available so that depth information can be recovered.

**Image Segmentation**

Our goal is to automatically determine the regions in the image where a particular food is located. Once a food item is segmented, we will identify the food item and then estimate how much food is present in the image.

Our image segmentation method is a two step process. In the first step the image is converted to a grayscale image and then thresholded with a threshold of 127 to form a binary image. It was determined empirically that the pixel values in the binary image corresponding to the plate had values of 255. For segmenting the food items on the plate, the binary image was searched in 8-point connected neighbors for the pixel value 0. Connected segments less than 1000 pixels were ignored because they correspond to the tablecloth (see Figure 3). In this step we used a fixed threshold. Thus, pixels corresponding to the food items might be considered as the plate. As a result, we need a second step of segmentation. The result of the first step of segmentation is shown in Figure 5(a).

In the second step, the image is first converted to the YCbCr color space. Using the chrominance components, Cb and Cr, the mean value of the histogram corresponding to the plate was found. Pixel locations which were not segmented during the first step were compared with the mean value of the color space histogram of the plate to identify potential food items. These pixels were labeled differently from that of the plate. Then 8-point connected neighbors for the labeled pixels were searched to segment the food items. An example is shown in Figure 5(b), the food item, i.e. scrambled egg, which was not segmented in the first step is successfully segmented in the second step using the color space.

**Feature Extraction**

Two types of features were extracted from each segmented food region, namely color features and texture features. For color features, the average value of the pixel intensity along with the two color components were extracted. For texture features, we used Gabor filters to measure local texture properties in the frequency domain.

Gabor filters describe properties related to the local power spectrum of a signal and have been used for texture features [38]. A Gabor impulse response in the spatial domain consists of a sinusoidal plane wave of
some orientation and frequency, modulated by a two-dimensional Gaussian envelope and is given by:

\[
h(x, y) = \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right] \cos (2\pi U x + \varphi)
\]  

(1)

In our work, we used a Gabor filter-bank proposed in [39]. It is highly suitable for our use where the texture features are obtained by subjecting each image (or in our case each block) to a Gabor filtering operation in a window around each pixel and then estimate the mean and the standard deviation of the energy of the filtered image. A Gabor filter-bank consists of Gabor filters with Gaussians of several sizes modulated by sinusoidal plane waves of different orientations from the same Gabor-root filter as defined in Equation (1), it can be represented as:

\[
g_{m,n}(x, y) = a^{-m} h(\tilde{x}, \tilde{y}), \quad a > 1
\]  

(2)

where \( \tilde{x} = a^{-m}(x \cos \theta + y \sin \theta) \), \( \tilde{y} = a^{-m}(-x \sin \theta + y \cos \theta) \), \( \theta = n\pi/K \) (\( K \) = total orientation, \( n = 0, 1, ..., K-1 \), and \( m = 0, 1, ..., S-1 \)), and \( h(\cdot, \cdot) \) is defined in Equation (1). Given an image \( I_E(r, c) \) of size \( H \times W \), the discrete Gabor filtered output is given by a 2D convolution:

\[
I_{g_{m,n}}(r, c) = \sum_{s, t} I_E(r - s, c - t) g_{m,n}^s(s, t),
\]  

(3)

As a result of this convolution, the energy of the filtered image is obtained and then the mean and standard deviation are estimated and used as features. In our implementation, we divided each segmented food item in to 64 × 64 non-overlapped blocks and used Gabor filters on each block. We used the following parameters: 4 scales \( (S=4) \), and 6 orientations \( (K=6) \).

**Classification**

Once the food items are segmented and their features are extracted, the next step is to identify the food items using statistical pattern recognition techniques [40, 41]. For classification of the food item, we used a support vector machine (SVM) [42–44]. The feature vectors used for the SVM contain 51 values, 48 texture features and 3 color features. The feature vectors for the training images (which contain only one food item in the image) were extracted and a training model was generated using the SVM. A set of 17 images were used as training images and each food item was given an unique label.

We created a database of images using a digital camera and plastic food replicas in our laboratory. The images were acquired using specific conditions, such that the foods were placed on a white plate on a checkerboard (black and white) patterned tablecloth. The tablecloth was used as a fiducial mark for estimating the dimensions and area of the food item. The white plates were used to assist the segmentation of the food items. The training images used were taken with only one food item and the testing image contained 2 or 3 food items. The database consists of 50 images, 17 images were used for training and 33 images were used for testing. The average classification results indicated a 93.745% accuracy when 17 images were used as training images and 14 images containing 32 food items were used as test images. Examples of correctly classified objects and misclassified objects are shown in Figure 6.

**Volume Estimation**

Based on the segmentation and reference size estimation, we need to determine the volume of food consumed in \( cm^3 \). For many foods this will not be possible from one image of the meal. We will explore several approaches in the future including the use of multiple images and computer visualization methods using 3D shape reconstruction techniques. We will generate a similar shaped object as the food item as a reference and composite it over the food item in the image and ask the user to adjust the shape to be smaller or larger to increase our accuracy in volume estimation. This involves the development of a combination of image processing and graphical rendering techniques to correctly position and size our estimated volume shape. The user simply adjusts a slider bar and confirms the choice of the size. This interactive user adjustment may become only an option in the final deployed system as our estimation algorithms increase in accuracy.

**Estimating Food Consumed**
Figure 6. Example of classified food items (a) All food items are successfully classified using a SVM, (b) Some food items are misclassified by the SVM, i.e. beef roast is misclassified as steak.

By taking digital images before and after the meal is eaten we will develop methods to estimate the amount of food consumed. Once we have determined the volume in $\text{cm}^3$ of each food item, this information will be combined with the portion code and portion weight in the USDA Food and Nutrient Database for Dietary Studies (FNDDS) to determine the gram weight of the food. For example, for homemade chocolate pudding, the food code is 13210220 in FNDDS. The portion code, 10205, associated with the chocolate pudding is 1 cup. The portion weight is 261 grams and 1 US cup is known to be $236.588237\text{cm}^3$. If the portion is calculated to be $100\text{cm}^3$, then the gram weight of the portion code and the $\text{cm}^3$ of the portion code are used to solve for the gram weight of the portion size. In this case the gram weight would be $110.318248$ grams. Once the gram weight is determined the nutritional information in the FNDDS for each food item can be used to determine the final energy and nutritional content of the consumed food. The FNDDS does not have built in volume measures so each portion code and portion weight will need to have an additional cubic centimeter field added.

4. CONCLUSIONS AND FUTURE WORK

Assessment of dietary intake is fraught with uncertainties. The goal of our work is to improve assessment of dietary intake through the further development of a novel and more precise way to measure food intake. We aim to establish a system, using a mobile device, to automatically determine what food was consumed at a meal by acquiring images of the food before and after the meal. Mobile computing devices provide a unique vehicle for collecting dietary information that reduces burden on record keepers.

Our initial effort has focused on identifying food items in an image using image analysis techniques. In our current method, we use simple 2D images for identification and quantification of food consumed. We are looking to create 3D food models to assist in pattern matching and to improve our estimated sizes of the food consumed.

REFERENCES


