Published in final edited form as:

J Am Diet Assoc. 2010 January ; 110(1): 42. doi:10.1016/j.jada.2009.10.011.

Automatic Food Documentation and Volume Computation using Digital Imaging and Electronic Transmission

Rick Weiss, MS,

President, Viocare, Inc., 145 Witherspoon Street, Princeton, NJ 08542, Phone: 609-497-4600- x10, Fax: 609-497-0660

Phyllis J. Stumbo, Ph.D., and

Research Nutritionist, University of Iowa, 157 Medical Research Facility, Iowa City, IA 52242, Phone: 319-384-9746, Fax: 319-384-8325

Ajay Divakaran, Ph.D

Technical Manager, Vision and Multi-Sensor Systems, Sarnoff Corporation, 201 Washington Road, PO Box 5300, Princeton, NJ 08543, Phone: 609-734-2204, Fax: 609-734-2662

Rick Weiss: weiss@viocare.com; Phyllis J. Stumbo: phyllis-stumbo@uiowa.edu; Ajay Divakaran: adivakaran@sarnoff.com

Keywords

Dietary assessment; speech recognition; image analysis; mobile phone; food intake; software

Introduction

Capturing accurate food intake data from participants enrolled in nutrition studies is essential for understanding relationships between diet and chronic disease (1). Numerous methods are employed to assess dietary intake such as food records, 24-hour recalls, or food frequency questionnaires. While each of these techniques is valuable, the error associated with each is unique. The food record requires a motivated participant, is tedious for some, places attention on the act of eating thus altering intake and is difficult for subjects with low literacy skills (2). Interviewing subjects about the previous day's intake avoids the reactivity involved when recording current intake, but also requires the individual reporting intake to have good recall skills, knowledge of food names and ability to estimate amounts eaten; and requires a well-trained interviewer which makes this a costly process (2,3). Food frequency questionnaires are limited by food lists and lack of detail regarding food preparation, and require respondents to summarize past intake over many months or the past year. Such instruments are known to contain significant measurement error (4). While all these methods provide valuable information about dietary intake, improving methodology even modestly would advance our knowledge about the influence of food intake on health.

FIVR (Food Intake Visual and voice Recognizer), a subproject of the Genes, Environment, and Health Initiative from the National Institutes of Health (RFA-CA-07-032 at www.gei.nih.gov/index.asp), is designed to use new digital photographing technology to

Correspondence to: Rick Weiss, weiss@viocare.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain

reduce measurement error associated with a food record. The intent is to create a tool that would both increase accuracy of intake records and reduce the recording burden for respondents. Using a mobile phone with a camera (Figure 1), the participant will photograph foods both before and after eating. In this way initial portion size is recorded as well as portions left uneaten. The photographs would be used to identify both the types and amounts of foods consumed. This paper briefly describes the technology and techniques involved.

Creating sufficiently detailed images

Capturing images of meals using a mobile phone presents its own unique challenges. Identifying foods from a picture requires a clear image; the automatic calculation of the amount eaten (volume) requires three or more clear images to be taken by the mobile phone user. Since a single image will not support estimation of food volume, rather 3-dimensional objects must be viewed at more than one angle (5,6). The three images in Figure 2 are captured from 3 slightly different angles. A calibration object is also required in the images for determination of 3-dimensional size (see Figure 2). The calibration object (fiducial marker) included in the images in Figure 2 is a card with black and white squares of known size. However, a standard credit card can be used to establish the relationship between size in image pixels and actual size of the object in milliliters. Images are also required before and after the meal is eaten to document the volume of food consumed.

Quality of the image hinges on several factors including resolution (roughly indicated by number of pixels per image). Higher resolution (more pixels per image) creates larger files, which makes transferring images slower and more subject to failure, thus testing and refinement of the image details is integral to developing a successful system. Camera focus is critical since the best volumetric estimation is obtained when the three images are in focus and taken with the plate at the same distance from the camera. With fixed focus cameras, the images will be blurred if not taken at the right distance (which is often too great). With auto-focus cameras, the focusing is assured but the distance still must be maintained by the user. Ways to adjust the image to correct for small variations in distance are still being explored.

Capturing intake data in real time

Using a mobile phone to document intake has several advantages. One worth mentioning is the ability to remind the user to provide intake information. The phone can be configured to issue queries about forgotten reports when the lag time between meals is greater than expected. The phone can also issue queries about what was consumed, asking the participant to describe the food intake orally, to be matched with the images when they are evaluated. Voice recognition software will be utilized to automate the food list capture and support queries to clarify food details. Timeliness is critical for improving accuracy of dietary reporting. Within a reasonable time the food scene should be translated into a written list of foods and amounts consumed, and the food list sent to the user for verification.

The final step will be matching the foods consumed to a food composition database. The USDA Food and Nutrient Database for Dietary Studies (FNDDS) (7) is the starting point used by FIVR. This database is used to evaluate the NHANES food intake studies, thus it broadly represents foods consumed in the United States. Each food will then be matched to the database of food names to identify a match. A human interface will resolve conflicts when the automatic program fails to find a match between spoken word and database. One important detail that is lacking in this automated system is the density of the foods. Density is required to convert volumes determined from images to weight, the unit of measure used in most databases. Most foods on the FNDDS provide the two components needed to calculate density (volume and weight), however some entries will require volume determinations to convert volume to weight.

Higher level databases such as the USDA MyPyramid Equivalents Database (8) can be incorporated so intake of whole foods (in addition to nutrients) can be evaluated.

Conclusion

Since phone signals are routinely directed through a central service, the phone can be configured to forward all image data to a server managed by the nutrition researcher or service provider thus creating a permanent record available in real-time for monitoring data collection and clarifying ambiguities while the information is fresh in the memory of respondents. Real-time processing of dietary intake data has only been available to researchers in metabolic ward studies, but technologies are now available to support real-time dietary assessment in the field. FIVR will use images to identify food items and estimate food volume, automatically translate food intake to nutrients and other components and classify foods consumed by major groups. FIVR is being developed to improve accuracy in research. However the precision and timeliness of the method will be equally valuable in clinical practice. This technology can lead the way to new services that will advance the study and practice of dietetics.

References

1. Kipnis V, Medthune D, Freedman L, Bingham S, Day NE, Riboli E, Ferrari P, Carroll RJ. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Pub Health Nutr* 2002;5:915–923. [PubMed: 12633516]
2. Tucker KL. Assessment of usual dietary intake in population studies of gene-diet interaction. *Nutr Metab and Cardio Dis* 2007;17:74–81.
3. Yanek LR, Moy TF, Becker DM. Comparison of food frequency and dietary recall methods in African American women. *J Am Dietet Assoc* 2001;101:1361–1364.
4. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158(1):14–21. [PubMed: 12835281]
5. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*. New York: Cambridge University Press; 2000.
6. Nister D. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2004;26(6):756–770. [PubMed: 18579936]
7. Bodner-Montville J, Ahuja JKC, Ingwersen LA, Haggerty ES, Enns CW, Perloff BP. USDA Food and Nutrient Database for Dietary Studies: Released on the web. *J Food Comp and Anal* 2006;19:S100–S107.
8. Bowman, SA.; Friday, JE.; Moshfegh, A. Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture; Beltsville, MD: 2008. MyPyramid Equivalents Database, 2.0 for USDA Survey Foods, 2003–2004 [Online] Food Surveys Research Group. Available at: <http://www.ars.usda.gov/ba/bhnrc/fsrg>

Figure 1.
Typical Mobile Phone Interface showing (a). operator instruction screen, (b) menu of activities available and (c) camera poised to record meal.

Figure 2.
Three images captured by moving the camera using the FIVR mobile phone system.