2012 IEEE International Conference on Multimedia and Expo Workshops

A New Texture Feature for Improved Food Recognition Accuracy in a Mobile Phone Based Dietary Assessment System

Md Hafizur Rahman[a], M. R Pickering[a], D. Kerr[b]
[a]School of Engineering and Information Technology
The University of New South Wales
Canberra, Australia

[b]School of Public Health
Curtin Unversity
Perth, Australia

C. J. Boushey[c], E. J. Delp[d]
[c]University of Hawaii Cancer Center
University of Hawaii
Hawaii, USA

[d]School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA

***Abstract*—Poor diet is one of the key determinants of an individual's risk of developing chronic diseases. Assessing what people eat is fundamental to establishing the link between diet and disease. Food records are considered the best approach for assessing energy intake however paper-based food recording is cumbersome and often inaccurate. Researchers have begun to explore how mobile devices can be used to reduce the burden of recording nutritional intake. The integrated camera in a mobile phone can be used for capturing images of food consumed. These images are then processed to automatically identify the food items for record keeping purposes. In such systems, the accurate classification of food items in these images is vital to the success of such a system. In this paper we will present a new method for generating texture features from food images and demonstrate that this new feature provides greater food classification accuracy for a mobile phone based dietary assessment system.**

Keywords-Dietary assessment, object recognition, texture features, Gabor filters, scale invariance

I. INTRODUCTION

There is now convincing evidence that poor diet, in combination with physical inactivity are key determinants of an individual's risk of developing chronic diseases, such as obesity, cancer, cardiovascular disease or diabetes. Preventing disease through improving nutrition is a global health priority [1]. Approximately 30% of all cancers have been attributed to dietary factors [2]. The strongest evidence for diet increasing cancer risk is specifically with overweight and obesity, high consumption of alcoholic beverages, aflatoxins and fermented foods. A diet of at least 400 g per day of fruits and vegetables appears to decrease cancer risk. However, a key barrier to linking dietary exposure and disease is the ability to measure dietary factors, including intake of food groups such as fruits and vegetables, with specificity and precision [3].

Assessing what people eat is fundamental to establishing the link between diet and disease. However, it is now more challenging to do this as consumers have moved away from eating a traditional 'meat and 3-veg' meal at home to purchasing more take-away food and eating out [4, 5]. With this greater proportion of foods eaten away from home [6, 7], it is now becoming increasingly difficult for consumers to accurately assess how much they have eaten or the composition of their meals.

Food records are considered the best approach for assessing energy intake (kilojoules). With a paper-based food record, individuals are asked to record their food and fluid intake for between 3-7 days. This method requires literate and highly motivated people. Research has shown adolescents and young adults, who typically have unstructured eating habits and frequently snack, are the least likely to undertake food records [8].

With advances in technology it is now timely to explore how mobile devices can better capture food intake in real-time by potentially reducing the burden of the recording task to both the study participant and the researcher. The ready access of the majority of the population to mobile phones has opened up new opportunities for dietary assessment which are yet to be leveraged. Tufano et al. [9] in a review of eHealth (web and mobile phone) applications refers to this as 'technology convergence' in which real-time or near-real time multimedia communication capabilities can occur.

The integrated camera in the mobile phone can be used for capturing images of food consumed. These images would then be processed to automatically identify the food items for record keeping purposes and to provide feedback to a patient. Hence, the accurate classification of food items in these images would be vital to the success of such a system and remains an open research problem.

In automatic food classification systems, compact features are first extracted from images of known food items in a database. When a query image of an unknown food item is input into the system, these same features are extracted from the query image. The features from the query image are compared with the features from each image in the database and the best matching features are found. The unknown food is then identified as the food item from the database which produced the best match to the features from the query image. The features used in this process can be extracted using many different characteristics of an image e.g. colour, texture, location and orientation of edges.

978-0-7695-4729-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMEW.2012.79

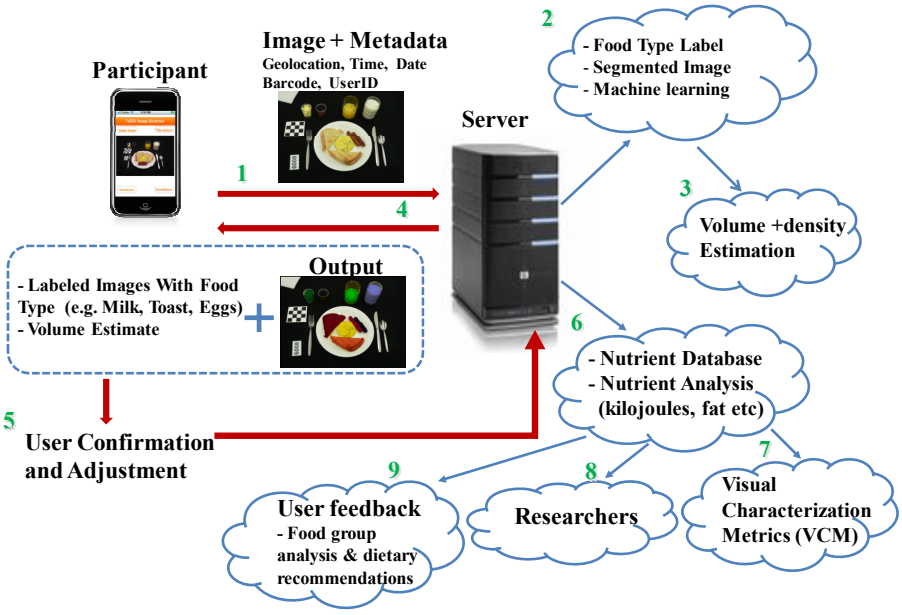

Figure 1. System architecture for the TADA dietary assessment system (http://www.tadaproject.org).

Some examples of previous approaches for automatic food recognition include a system proposed by Shulin Yang et al. [10] for recognizing fast food such as hamburgers, pizza and tacos. Their approach used the relative spatial relationships of local features of the ingredients followed by a feature fusion method to classify food images captured under laboratory conditions. Hoashi et al. [11] proposed an automatic food image recognition system for Japanese style food categories by fusing various kinds of local image features including bag-of-features (BoF), color histogram, Gabor features and gradient histogram with Multiple Kernel Learning. As an alternative approach, Zhu et al. proposed a food recognition system using only global features derived from color and texture [12]. Bosch et al. [13] extended these approaches to use a combination of global and local features.

Food items are deformable objects and have significant variation in appearance. In practical situations, images of a food item may be captured at different scales and/or orientations to images of the same food item currently in the database. The existing food recognition systems [10–13] extract non-invariant texture and color features and hence often provide unacceptable performance in classifying rotated and/or scaled versions of the query food image. Consequently researchers have begun to address the problem of developing features that are invariant to the scale and rotation of the texture.

In this paper we concentrate on extracting features using the textures in an image. We will present a new method for generating scale and/or rotation invariant global texture features using the output of Gabor filter banks. We will demonstrate that this new texture feature provides greater food classification accuracy for a mobile phone based dietary assessment system.

II. The Mobile Phone Based Dietary Assessment System

The new texture feature proposed in this paper will be utilized in the mobile phone based dietary assessment system that has been developed as part of the Technology Assisted Dietary Assessment (TADA) project [14,15] The system architecture for this system is shown in Figure 1. The users of the system are given a mobile phone with a built-in camera, network connectivity, and integrated image analysis and visualization tools to allow their food and beverage intake to be recorded. The process starts with the user sending the food image and contextual metadata (date, time, geolocation, etc.) to the server via the data network (step 1). The user places a fiducial marker in each image that allows the images to be spatially and colour calibrated. Food identification and volume estimation are performed on the server (steps 2 and 3). The results of step 2 and 3 are sent back to the user to confirm and/or adjust this information, if necessary (steps 4 and 5). Once the server obtains the user confirmation, food information stored in a database at the server is used for calculating the nutrient information (step 6). The database contains information on the most commonly consumed foods, their nutrient values, and weights and densities for typical food portions. The food database uses image based features, referred to as Visual Characterization Metrics (VCMs) (step 7), to index the nutrient information. Finally, these results are sent to the researcher for further analysis (step 8) and future developments will incorporate

user feedback including food group analysis and dietary recommendations (step 9).

III. Texture Feature Generation Using Gabor Filters

In the conventional approach to generating texture features, the image is filtered using a set of Gabor filters which pass spatial frequencies with different scales and orientations. Figure 2 (a) shows the impulse responses for part of a typical set of Gabor filters. For each pixel, the outputs of the complete set of filters can be combined into a two-dimensional feature in the scale-rotation space as shown in Figure 2 (b).

Researchers have begun to address the problem of developing features that are invariant to the scale and rotation of the texture. Han and Ma [16] addressed the problem of scale and rotation invariance by projecting the two-dimensional feature vector onto the scale dimension and the rotation dimension to produce two one-dimensional feature vectors. However some information about the original filter outputs is lost when projecting the features and this allows some ambiguity in the matching process.

Li et al. [17] proposed a method where scale and rotation invariance were achieved by calculating multiple scale energies of the Gabor filtered image. The optimal matching scales of the Gabor texture descriptor were chosen using a matched filter approach along the scale dimension. The main drawback of this method is the requirement to perform multiple matches in the scale dimension.

Lo et al. [18] proposed a scale and rotation invariant (SRI) feature generation method using the Double Dyadic Dual-Tree Complex Wavelet Transform (D3T-CWT). However their proposed feature generation method can also be implemented with Gabor filter outputs. In this approach their feature was generated by applying a two-dimensional FFT to the feature in scale-rotation space and then using the magnitudes of the output of this FFT as the scale and rotation invariant feature. However as only the magnitudes of the FFT outputs are used some information about the original filter outputs is lost.

IV. Invariant Texture Feature Generation

For a given image $I(x,y)$, its Gabor wavelet transform is given by the convolution:

$$G(m,n,x,y) = \sum_{s \in S} \sum_{t \in T} I(x-s, y-t) g^*_{mn}(s,t) \qquad (1)$$

where g^*_{mn} represents the complex conjugate of the coefficients of the Gabor filter for scale m and rotation n, and S and T denote the region of support of the filter. For an image region, the standard Gabor texture feature is taken as the mean and standard deviation of the filter outputs at each scale and rotation and is given by:

$$F = \left[\mu_{00}\ \sigma_{00} \cdots \mu_{(M-1)(N-1)}\ \sigma_{(M-1)(N-1)} \right] \qquad (2)$$

where M is the number of scales and N is the number of rotations used in the Gabor filter bank.

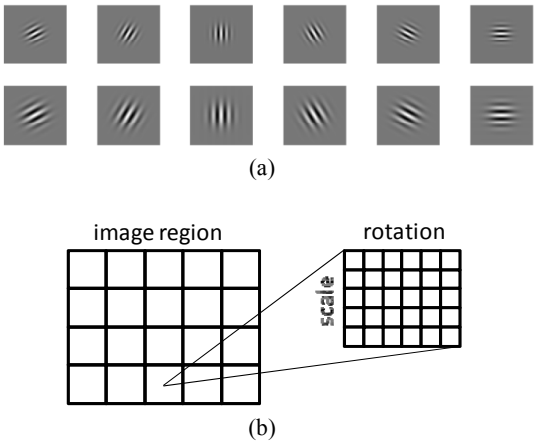

Figure 2. (a) The real part of a set of Gabor filters for 2 scales and 6 orientations (b) Two dimensional feature vectors are formed from the filter outputs at each pixel of the image region.

In our proposed approaches for providing scale invariance we apply a Gaussian window to weight the filter outputs. The coefficients of the window used in our algorithms are computed using the following equation.

$$W(k) = e^{-(\alpha k / N)^2} \qquad (3)$$

where $-N \le 2k \le N$, the window length is $L = N + 1$ and the width of the window is inversely proportional to α.

A. Rotation Invariant Texture Features

Our proposed algorithm to produce rotation invariant texture features is described in pseudo-code as follows:

1. Filter the input image $I(x,y)$ to give filter outputs for M scales and N orientations using (1) and store in the 4D array $G(m,n,x,y)$.
2. For each pixel position (x,y) and each scale m extract a vector V of length N.
3. Apply a circular shift on V to shift the maximum value to be the first element in V.
4. For each pixel position (x,y) and each scale m insert the vector V to give the new 4D array $G^{RI}(m,n,x,y)$.
5. The RI Gabor feature vector is calculated from the mean and standard deviations for scale m and orientation n in G^{RI} and is given by:

$$F^{RI} = [\mu^{RI}_{00}\ \sigma^{RI}_{00} \cdots \mu^{RI}_{(M-1)(N-1)}\ \sigma^{RI}_{(M-1)(N-1)}]$$

B. Scale Invariant Texture Features

Our proposed algorithm to produce scale invariant texture features is described in pseudo-code as follows:

1. Filter the input image $I(x,y)$ to give filter outputs for M scales and N orientations using (1) and store the magnitude of the complex values in the 4D array $G(m,n,x,y)$.

2. For each pixel position (x,y) and each orientation n extract a vector V of length M.
3. Multiply V by the Gaussian window W of length M and store in the vector U.
4. Apply a circular shift on U to shift the maximum value to be the first element in U.
5. For each pixel position (x,y) and each orientation n insert the vector U to give the new 4D array $G^{SI}(m,n,x,y)$.
6. The SI Gabor feature vector is calculated from the mean and standard deviations for scale m and orientation n in G^{SI} and is given by:

$$F^{SI} = [\mu_{00}^{SI}\ \sigma_{00}^{SI} \cdots \mu_{(M-1)(N-1)}^{SI}\ \sigma_{(M-1)(N-1)}^{SI}]$$

The application of the Gaussian window in step 3 of our approach is designed to decrease the impact of very high and very low spatial frequencies in the feature vector.

C. Scale and Rotation Invariant Texture Features

We also propose a method for extracting scale and rotation invariant features by merging the individual processes of generating scale invariant and rotation invariant features, described above. The details of the algorithm are described in pseudo code as follows:

1. Filter the input image I(x,y) to give filter outputs for M scales and N orientations using (1) and store in the 4D array G(m,n,x,y).
2. Generate scale invariant texture features as described in section 3.2 and store in the 4D array GSI(m,n,x,y).
3. Apply the process to generate rotation invariant features on GSI and store the results in the 4D array GSRI(m,n,x,y).
4. The SRI Gabor feature vector is calculated from the mean and standard deviations for scale m and orientation n in GSRI and is given by:

$$F^{SRI} = [\mu_{00}^{SRI}\ \sigma_{00}^{SRI} \cdots \mu_{(M-1)(N-1)}^{SRI}\ \sigma_{(M-1)(N-1)}^{SRI}]$$

V. The Food Image Databases

To evaluate the classification performance of our proposed scale and rotation invariant feature extraction methods we performed experiments using a database containing images from 209 food categories such as apple, bread, pizza, pasta, burger, potato chips etc. Each food item was placed on a white plate and an Apple iPhone was used to capture an image of the food item. Each image had a resolution of 1200 × 1600 pixels. Figure 3 shows some images from the food database. The original images contained some additional items e.g. plate, checker board pattern etc. In order to make the retrieval process more efficient, we segment each image and isolate the food region from the background. Figure 4 shows an example of the segmentation of a pink lady apple.

To assess the specific performance of each of the feature vectors tested, we used the images in the database to generate the following three databases:

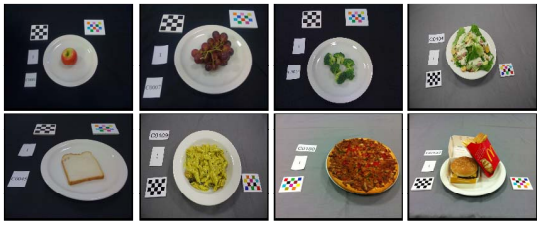

Figure 3. Some items from our food database.

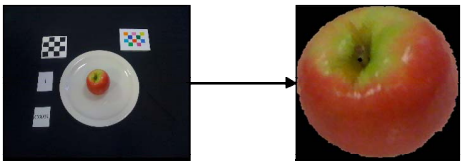

Figure 4. Isolation from the background by segmentation

A. The Rotation Invariance (RI) Database

In this database, each image was rotated at seven different rotation angles: 0°, 30°, 60°, 90°, 120°, 150° and 200° and thus, we have 1463 sample images (209 ×7 = 1463) with 7 images in each class.

B. The Scale Invariance (SI) Database

For this database we limited the range of different scales for each food item since even humans will perceive a different pattern when the scaling is above a certain level. Hence the 111 images were scaled with factors ranging from 0.7 to 1.4 at intervals of 0.1. Thus, we have 1672 sample images (209 × 8 = 1672) with 8 images in each class.

C. The Scale and Rotation Invariance (SRI) Database

For this database, joint scaling and rotation transforms were performed on the images. Each image was scaled by 0.7 to 1.4 with 0.1 intervals and rotated at seven different rotation angles: 0°, 30°, 60°, 90°, 120°, 150° and 200° Thus, we have 11,704 sample images (209×8×7 = 117,04) with 7 orientations and 8 scales providing 56 images in each class.

VI. Performance Evaluation

In the following experiments, each image in the database was used as a query image and the relevant images were the other scaled and/or rotated images of the same food item. The Canberra distance [19] was used to determine the similarity between feature vectors. The number of scales and rotations in the texture feature was set to 5 and 6 respectively and the value of α that controls the width of the Gaussian window in the scale dimension was set to 1.2. This value of α was determined experimentally to give the best results.

A. Experiment 1: Image retrieval

In this experiment the new texture feature was evaluated for its general image retrieval ability. The standard approach for evaluating image retrieval performance is through

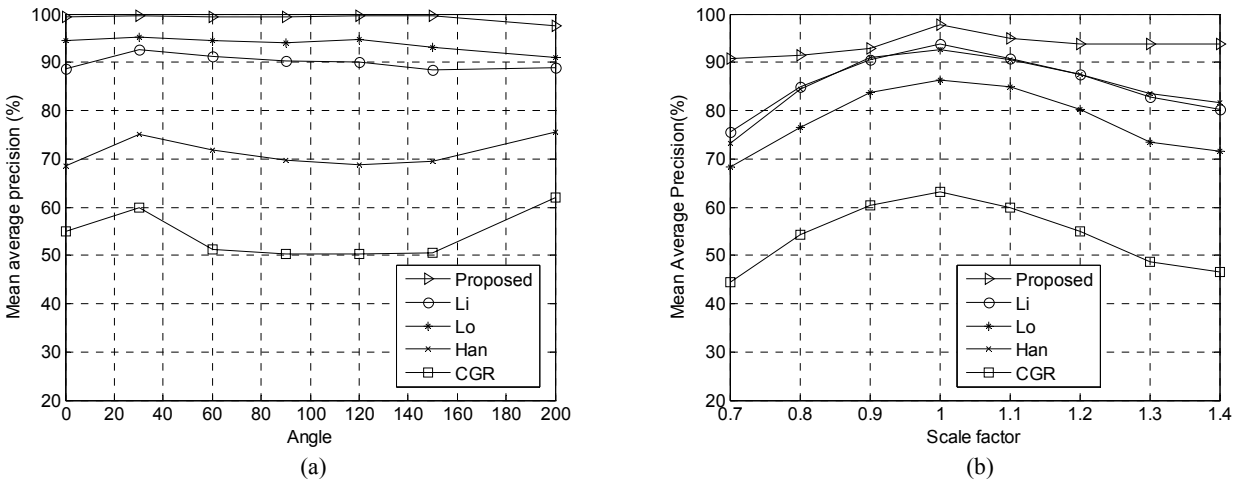

Figure 6. (a) MAP versus rotation angle of the query image. (b) MAP versus scale factor of the query image.

computation of Precision – Recall (P-R) curves [18]. Precision at n for n retrieved images is defined as the number of relevant images retrieved divided by n. Recall is the number of relevant images retrieved divided by the total number of relevant documents in the database. P-R graphs show the variation of precision at n for recall values between 0 and 1. Perfect retrieval is achieved when precision is 1 for all values of recall. The mean average precision (MAP) effectively measures the area under a precision-recall curve and is often used to provide a single figure of merit for retrieval performance [20]. The performance of our proposed method was compared with four other feature generation methods: the conventional Gabor representation (CGR) and the methods of Han [16], Li [17] and Lo [18].

Figure 5 (a) shows the MAP for each feature generation method as the rotation angle of the query images is varied. Notice that the retrieval performance for the proposed approach is almost 100% for all rotation angles and is significantly better than for other previously proposed approaches. Figure 5 (b) shows the MAP as the scale factor of the query images is varied. These curves also show significantly better performance for the proposed approach. In particular notice how the MAP values remain above 90% for the extreme scale factor values of 0.7 and 1.4. Table 1 shows the overall MAP values for the five methods tested for the SI, RI and SRI databases.

TABLE I. MEAN AVERAGE PRECISION (%) FOR SCALE AND ROTATION INVARIANT IMAGE RETRIEVAL.

	CGR	Han	Lo	Li	Proposed
RI	54.15	71.28	90.03	93.89	**99.46**
SI	58.70	73.14	85.13	83.21	93.70
SRI	55.87	70.74	80.28	82.96	**94.18**

B. Experiment 2: Semi-Automatic Food Classification

In this experiment the new feature was evaluated for its ability to classify individual food items. In the TADA system, the classification system provides the user with a candidate list of possible food labels for each food in the query image. This list is displayed on the screen of the mobile phone when the user selects a food item in the image. The user then chooses the correct food label out of the list displayed. This means that the food classification system only needs to provide the correct food label in the list displayed to the user. In this experiment the new texture feature was evaluate for its performance in providing the correct label in a list of top ranked labels of sizes from one (fully automatic) to five. Table 2 shows the percentage of times that the correct label was displayed in the candidate list that would be displayed to the user.

TABLE II. SEMI-AUTOMATIC FOOD CLASSIFICATION PERFORMANCE (%).

	Length of candidate list				
	1	**2**	**3**	**4**	**5**
RI	99.46	99.64	99.82	99.91	100.00
SI	94.93	96.38	98.52	99.59	99.95
SRI	95.93	97.38	98.52	99.28	99.85

C. Experiment 3: Automatic Food-Group Classification

In this experiment the classification problem was simplified to classifying the food items in the query image into five broad categories: Breads, Cereal, Fruits, Vegetables and Fast Food. This classification task is intended to be used to generate automatic text messages to users. For example when the user does not eat the required number of serves of fruit and vegetables in a day or is eating too many fast food meals. Table 2 shows the percentage of times that the query image was identified as belonging to the correct broad food category.

TABLE III. AUTOMATIC FOOD GROUP CLASSIFICATION PERFORMANCE (%).

	Food Group Category				
	Bread	**Cereal**	**Veg**	**Fruit**	**Fast**
RI	100.00	100.00	99.95	99.88	99.85
SI	98.59	99.87	97.12	97.45	94.33
SRI	99.75	99.91	97.82	98.52	95.89

VII. CONCLUSIONS

A new method to extract features that are invariant to the scale and rotation of the texture in an image was presented. Experimental results show that the new features allow significantly better food classification accuracy than existing approaches. The large increase in retrieval performance between the proposed approach and the conventional non-invariant approach highlights the need for such invariant features in any practical food recognition system.

For individual food item recognition we have demonstrated that, when using the proposed texture feature, the correct food label occurs in the list of the top five ranked labels almost 100% of the time. Our results also show that when using only the proposed texture feature, food items in our database can be classified into the correct broad categories with almost 100 % accuracy.

REFERENCES

[1] World Health Organization, World Health Assembly resolution WHA57.17 – "Global strategy on diet, physical activity and health." 2003, World Health Organization: Geneva.

[2] World Health Organization, WHO Technical Report Series 916 – "Diet, nutrition and the prevention of chronic diseases," R.o.a.J.W.F.E. Consultation, Editor. 2003: Geneva.

[3] B. M. Margetts, and M. Nelson, "Overview of the principles of nutritional epidemiology," in Design concepts in nutritional epidemiology, B.M. Margetts and M. Nelson, Editors. 1997, Oxford University Press: Oxford ; New York. p. 3-38.

[4] Australian Bureau of Statistics, "Cafes, Restaurants and Catering Services, Australia, 2006-07." 2008, Australian Bureau of Statistics, Canberra.

[5] Australian Government Department of Agriculture Fisheries and Forestry, "FOODmap. A comparative analysis of Australian food distribution channels," Commonwealth of Australia, Editor. 2006.

[6] J. Dixon, et al., "The health equity dimensions of urban food systems," J Urban Health, 2007. 84(3 Suppl): p. i118-29.

[7] C. Burns, et al., "Foods prepared outside the home: association with selected nutrients and body mass index in adult Australians," Public Health Nutr, 2002. 5(3): p. 441-8.

[8] C. J. Boushey, et al., "Use of technology in children's dietary assessment," European Journal of Clinical Nutrition, 2009. 63: p. S50-S57.

[9] J. T. Tufano and B.T. Karras, "Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends," J Med Internet Res, 2005. 7(5): p. e58.

[10] Y. Shulin, et al., "Food recognition using statistics of pairwise local features," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2249-2256.

[11] H. Hoashi, T. Joutou, and K. Yanai, "Image Recognition of 85 Food Categories by Feature Fusion," in Multimedia (ISM), 2010 IEEE International Symposium on, 2010, pp. 296-301.

[12] F. Zhu, et al., "Segmentation Assisted Food Classification for Dietary Assessment," in Proceedings of the IS&T/SPIE Conference on Computational Imaging IX, Vol. 7873, San Francisco Airport, California, January, 2011.

[13] M. Bosch, et al., "Combining Global and Local Features for Food Identification and Dietary Assessment," in Proceedings of the International Conference on Image Processing (ICIP) Brussels, Belgium, 2011.

[14] Schap TE, Six BL, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. 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 Nutr 2011:1-8.

[15] Mariappan A, Bosch Ruiz M, Zhu F, et al. Personal Dietary Assessment Using Mobile Devices. In: Proceedings of the IS&T/SPIE Conference on Computational Imaging VII; 2009; San Jose; 2009.

[16] J. Han and K.-K. Ma, "Rotation-invariant and scale-invariant Gabor features for texture image retrieval," Image and Vision Computing, vol. 25, pp. 1474-1481, 2007.

[17] Z. Li, et al., "Scale and rotation invariant Gabor texture descriptor for texture classification," in Proceedings of the SPIE, Conference on Visual Communications and Image Processing 2010, Huangshan, China, 2010.

[18] E. H. S. Lo, et al., "Scale and rotation invariant texture features from the dual-tree complex wavelet transform," in Image Processing, 2004. ICIP '04. 2004 International Conference on, 2004, pp. 227-230 Vol. 1.

[19] M. Kokare, B. N. Chatterji and P. K. Biswas, "Comparison of Similarity Metrics for Texture Image Retrieval," Proc. IEEE TENCON Conf., Bangalore, India, 2003, pp. 571-575

[20] C. D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, Cambridge, England, 2008