Using graph cut segmentation for food calorie measurement

CONFERENCE PAPER · JUNE 2014

DOI: 10.1109/MeMeA.2014.6860137

Using Graph Cut Segmentation for Food Calorie Measurement

Parisa Pouladzadeh[1], Shervin Shirmohammadi[1, 2], Abdulsalam Yassine[1]
[1]Distributed and Collaborative Virtual Environment Research Laboratory
University of Ottawa, Canada
Email: {ppouladzadeh, shervin, ayassine}@discover.uottawa.ca
[2] Colleges of Engineering and Natural Sciences, Istanbul Şehir University, Turkey

ABSTRACT

Calorie measurement systems that run on smart phones allow the user to take a picture of the food and measure the number of calories automatically. In order to identify the food accurately in such systems, image segmentation, which partitions an image into different regions, plays an important role. In this paper, we present the implementation of Graph cut segmentation as a means of improving the accuracy of our food classification and recognition system. Graph cut based method is well-known to be efficient, robust, and capable of finding the best contour of objects in an image, suggesting it to be a good method for separating food portions in a food image for calorie measurement. In this paper, we provide the analysis of the Graph cut algorithm as applied to food recognition. We also perform a number of experiments where we used results from the segmentation phase to the Support Vector Machine (SVM) classification model. The results show an improvement in the accuracy of food recognition, especially mixed food where accuracy increases by 15% compared to our previous work [10].

Index Terms—**Food recognition, Segmentation, Graph cut, Classification.**

I. INTRODUCTION

Recently, people are becoming used to sedentary lifestyle since they can be fully consumed by busy schedules at work and at home. Not only they cannot find time to exercise, but also they are involuntarily carried away from being cognizant about their food consumption. Many studies are relating such lifestyle to the ever increasing number of obese people worldwide [1]. There is plentiful evidence that obesity increases the risk of serious diseases such as diabetes, blood pressure, hypertension etc. Similarly, there is abundant evidence that diet-related health conditions are among the leading causes of death and that eating healthily diminishes those risk conditions [1].

While people in general understand the links between diet and health, knowledge alone cannot be the solution. The main issue is how to assist people to become aware of their calorie consumption and diet choices in order to control obesity rates. In most cases, people are oblivious about measuring or controlling their calorie intake due to the lack of nutrition knowledge or willpower. In this paper, we propose a calorie measurement system that runs on smart phones, which allows the user to take a picture of the food and measure the number of calories automatically. Our aim is to empower users by a convenient, intelligent and an accurate system that helps them become aware of their calorie intake. In order to identify the food accurately in our system, image segmentation, which partitions an image into different regions, plays an important role. In this paper, we present the implementation of graph cut segmentation as a means of improving the accuracy of our food classification and recognition system.

In our previous work, see, [2][3], we have shown an overview of our food detection and recognition method. We have used image processing and Support Vector Machine (SVM) methods for food recognition and calorie measurement. To increase the measurement accuracy of the calorie intake, it is necessary to improve the food classification accuracy of the system. In this paper, we apply Graph cut segmentation to our food recognition system. Graph cut segmentation is a perfect method for selecting regions and finding the best boundaries for each food portions [5][6]. To use graph cut segmentation, an important question will be: what should be the features of a good graph extracted from an image? The answer is threefold: First, it should be robust; i.e., if the image is somewhat distorted, the graph should not be deeply changed. In graph cut, each pixel of the image is mapped onto a vertex in a graph. Neighboring pixels are connected by weighted edges where the weight is determined based on a predefined energy function. In the normalized cut approach, the cut cost is determined by the fraction of the total edge connections to all the vertices in the graph [5]. Second, it should also have good algorithmic properties. This means that the graph, when drawn, is actually a symbolic representation of the image; for instance, the boundaries between the regions should match with the edges of the graph. Third, we would like to be able to rebuild an image from the graph and for this new image to be a good compression of the initial image. In other terms, the loss due to the extraction process should be minimal. In this paper, the above three properties

978-1-4799-2921-4/14/$31.00 ©2014 IEEE

are used in the design of our graph cut based food image segmentation method. We performed a number of experiments where we used results from the segmentation phase to the Support Vector Machine (SVM) classification model. The results show an improvement in the accuracy of the system in single food portions. More importantly, the results show a significant increase in recognizing mixed foods when comparing the results with [10].

The rest of the paper is organized as follows: In section II, we present the related work. Section III, gives our proposed system. In section IV we describe the experimental results. Finally, in section V, we conclude the paper and provide direction for future work.

II. RELATED WORK

This section presents some of the most common food intake measuring methods. The aim of this section is to provide the advantages and disadvantages of existing methods and the needs for alternative approaches.

In [10], a web application based system is proposed which detects whether the user has the habits considered as risk factors of obesity. The goal of this system is to inform users of their particular habits that influence their weight-gain. The application acquires and registers data about diet, exercise, sleep, and fat mass, by using a web application and health information sensors. In [11], the authors proposed a mobile game that helps motivate children to practice healthy eating habits by letting them take care of a virtual pet. The above two systems do not provide any measurement of calories of the consumed food and it is up to the user to keep a log of his/her behaviour. The major drawback of such systems is that it is inconvenience for the user and the learning process is troublesome. Also, the users must be adequately motivated before they change their behaviour.

To alleviate the aforementioned issues, researchers have been studying other convenient systems that use food images to provide users with information about their food consumption and to calculate their calorie intake, see, e.g. [12], [13], [14], [15] and [16]. In [12] the authors proposed a system that utilizes food images that are captured and stored by multiple users in a public Web service called FoodLog. Then, a dictionary dataset of 6512 images is formed including the calorie estimation. The images in the dictionary are used in dietary assessment approach. The accuracy in such approach for measuring calories is very low. In [13], a new 3D/2D model-to-image registration framework is presented for estimating food volume from a single-view 2D image containing a reference object. In this system, the food is segmented from the background image using morphological operations while the size of the food is estimated based on user-selected 3D shape model. In [14] and [15], a set of pictures is taken for before and after food consumption in order to recognize and classify the food and determine its size. In such method, the existence of a premeasured and predefined measurement pattern is used inside the images to translate the size in pixels of each portion. All these conditions can generate difficulties, which has been addressed by Martin et al. [16]. In [16], the authors proposed a system that captures the images and sends them to a research facility where the analysis and extraction are performed. The major disadvantage of such system is that it does not provide information to the users in real-time. There is a considerable delay in providing the information due to the offline processing of images.

In our calorie measurement system, we also use food images taken with the built-in camera of a smartphone. But, we go one step further in terms of providing accurate measurement of the calorie intake, by implementing a graph cut segmentation as a means of improving the accuracy of our food classification and recognition system. Furthermore, the processing of the images and the calorie measurement results are provided to the user instantly. The details of the proposed system are discussed in the next section.

III. PROPOSED SYSTEM

In this section, we discuss our proposed system in more details. Specifically, we provide analysis of the graph-based segmentation and the food classification method.

A detailed block diagram of the proposed system is shown in Figure 1.

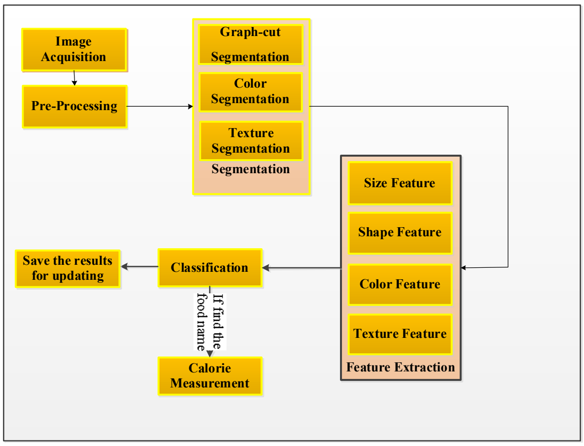

Figure 1.Proposed System

As the figure shows, at the early stage, images are taken by the user with a mobile device followed by a pre-processing step. Then, in the segmentation step, each image will be analyzed to extract various segments of the food portion. It is well-known that without having a good image segmentation mechanism, it is not possible to process the image appropriately. Hence, our attention in this paper is given to the segmentation step. To do so, we have used Graph-cut segmentation, color segmentation, and texture segmentation. Later in the paper, we show how these steps lead to an accurate food separation scheme. For each detected food portion, a feature extraction process has to be performed. In this step, various food features including size, shape, color and texture are extracted. The extracted features are then sent to the classification step where an SVM scheme is used for food portion identification. Finally, by estimating the area of the food portion and using nutritional tables, the caloric value of the food is calculated. In the following we will discuss more about segmentation and classification steps.

A. Segmentation

1) Graph-Based Segmentation

In this section, we introduce the graph cut method, which is used to establish the graph with the given image for segmentation. Let G=<V, E> be an undirected graph where V is a series of vertices and E is the graph edge which connects every two neighboring vertices. The vertex V is composed of two different kinds of nodes (vertices). The first kind of vertices is neighborhood nodes, which correspond to the pixels and the other kind of vertices are called terminal nodes, which consist of s (source) and t (sink). This type of graph is also called s-t graph where, in the image s node usually represents the object while t node denotes the background. In this graph, there are also two types of edges. The first type of edges is called n-links, which connects the neighboring pixels within the image. And the second type of edges is called t-links, which connects the terminal nodes with the neighborhood nodes. Each edge is assigned with a non-negative weight denoted as w_e, which is also named as cost. A cut is a subset of edges E, which can be denoted as C and expressed as $C \subset E$. The cost of the cut $|C|$ is the sum of the weights on edges C, which is expressed as follows.

$$|C| = \sum_{e \in C} W_e \quad (1)$$

A minimum cut is the cut that has the minimum cost called min-cut and it can be achieved by finding the maximum flow, which is verified in [5] to be equivalent to the max-flow. The max-flow/min-cut algorithm developed by can be used to get the minimum cut for the s-t graph. Thus, the graph is divided by this cut and the nodes are separated into two disjoint subsets S and T where $s \in S$, $t \in T$ and $S \cup T = V$. The two subsets correspond to the foreground and background in the image segmentation. An illustration of this graph is depicted in Figure 2.

Image segmentation can be regarded as pixel labeling problems. The label of the object (s-node) is set to be 1 while that of the background (t-node) is given to be 0 and this process can be achieved by minimizing the energy-function through minimum graph cut. In order to make the segmentation reasonable, the cut should be occurred at the boundary between the object and the background.

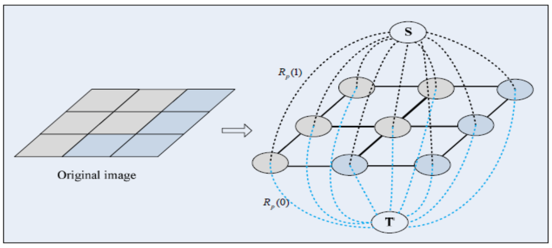

Figure 2.Illustration of s-t graph. The image pixels correspond to the neighbor nodes in the graph. The solid lines in the graph n-links and the dotted lines are t-links.

Namely, at the object boundary, the energy (cut) should be minimized. Let $L = \{l_1, l_2, l_3, \dots l_i, \dots l_p\}$, where p is the number of the pixels in the image and $l_i \in \{0,1\}$, thus, the set L is divided into 2 parts and the pixels labelled with 1 belong to object while others are grouped into background. The energy function is defined in equation (2), which can be minimized by the min-cut in the s-t graph [6].

$$E(L) = \propto R(L) + B(L) \quad (2)$$

where, R(L) is called regional term, which incorporates the regional information into the segmentation and B(L) is called boundary term, which incorporates the boundary constraint into segmentation, E(L) is the relative importance factor between regional and boundary term. When $\propto$ is set to be 0, it means that the regional information is ignored and only considering the boundary information. In the energy function in equation (2), the regional term is defined as follows:

$$R(L) = \sum_{p \in P} R_p(l_p) \quad (3)$$

Where, $R_p(l_p)$is the penalty for assigning the label l_p to pixel p. The weight of $R_p(l_p)$can be obtained by comparing the intensity of pixel p with the given histogram (intensity model) of the object and background. The weight of the t-links is defined in the following equations [5][6].

$$R_p(1) = -lnPr(I_p|'obj') \quad (4)$$

$$R_p(0) = -lnPr(I_p|'bkg') \quad (5)$$

From equation (4), (5), we can see that when $\Pr(I_p|'obj')$ is larger than $\Pr(I_p|'bkg')$ will be smaller than $R_p(0)$. This means when the pixel is more likely to be the object, the penalty for grouping that pixel into object should be smaller which reduces the energy in equation (2). Thus, when all of the pixels are correctly separated into two subsets, the regional term is minimized. B (L) in equation (2) is the boundary term, which is defined in the following equation.

$$B(L) = \sum_{\{p,q\} \in N} B_{<p,q>} . \delta(I_p, I_q) \quad (6)$$

Where p, q is neighboring pixels and $\delta(I_p, I_q)$ is defined as:

$$\delta(I_p, I_q)=\begin{cases}1 \; if \; I_p = I_q \\ 0 \; if \; I_p \neq I_q\end{cases} \tag{7}$$

For the regional constraint, it can be interpreted as assigning labels $\mathrm{I_p}, \mathrm{I_q}$ to neighboring pixels. When the neighboring pixels have the same labels, the penalty is 0 which means that the regional term would only sum the penalty at the segmented boundary. For the term $\mathrm{B}_{<p,q>}$, it is defined to be a non-increasing function of $|\mathrm{I_p} - \mathrm{I_q}|$ as follows [7]:

$$B_{<p,q>} \propto exp(-\frac{(I_p - I_q)^2}{2\sigma^2}) \tag{8}$$

Where σ can be viewed as camera noise. When the intensity of two neighboring pixel is very similar, the penalty is very high. Otherwise, it is low. Thus, when the energy function obtains a minimum value, it is more likely occurred at the object boundary. The study in [6] shows that the minimized energy value can be computed by the min-cut through max-flow. Consequently, the minimum energy problem is converted into the graph cut problem. In order to get a reasonable segmentation result, the assignment of the weight in the s-t graph is very important. In [6], the weight of the s-t graph is given as following.

$$Weight = \begin{cases} B_{<p,q>} \; \{p,q\} \in \; Neiboring \; pixel \\ \propto . R_p(0) \qquad for \; edge \; \{p,s\} \\ \propto . R_p(1) \qquad for \; edge \; \{p,T\} \end{cases} \tag{9}$$

Equation (9) can also be explained as such, in the s-t graph, when the intensity of the pixel is inclined to be the object, the weight between this pixel and the s-node will be larger than that between the pixel and the t-node, which means that the cut is more likely occurred at the edge with smaller weight. For the neighboring pixels, when their intensity is very similar, the weight is very big, which is not likely to be separated by the cut. Thus, when the minimum cut is achieved from the s-t graph, the location of the cut is close to the object boundary. The implementation of the graph cut can be fulfilled by the max-flow/min-cut as described in [5], [6], [7]. In Figure 3, we illustrate the graph cut for a 3×3 image segmentation. The thickness of the edge denotes the magnitude of the weight.

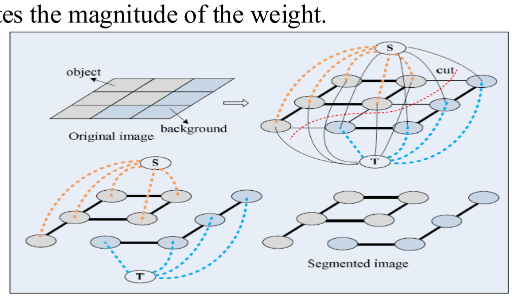

Figure 3 Illustration of graph cut for image segmentation.

2) Texture segmentation

For texture features, we used Gabor filters to measure local texture properties in the frequency domain. We used the Gabor filter-bank proposed in [4], which is highly suitable for our purpose where the texture features are obtained by subjecting each image to a Gabor filtering operation in a window around each pixel. We can then estimate the mean and the standard deviation of the energy of the filtered image. The size of the block is proportional to the size of the segment. 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. It is given by:

$$h(x,y) = -\exp\frac{1}{2}\left\{\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right\}\cos(2\pi U_x + \varphi) \tag{10}$$

Where U_x and φ are the frequency and phase of the sinusoidal plane wave along the z-axis (i.e. the 0°orientation), and $\sigma_x$and $\sigma_y$are the space constants of the Gaussian envelope along the z- and y-axis, respectively.

A Gabor filter-bank consists of Gabor filters with Gaussian kernel function 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:

$$g_{m,n}(x,y) = a^{-m} h(x',y'), a > 1 \tag{11}$$

Where:

$$x' = xcos\theta + ysin\theta$$

$$y' = -xsin\theta + ycos\theta$$

$$\theta = \frac{n\pi}{k}(k = \; total \; orientation, n = 0,1,\ldots,k - 1. m = 0,1,\ldots s - 1)$$

Give an Image $I_E(r,c)$ of size HxW , the discrete Gabor filtered output is given by a 2D convolution:

$$I_g(r,c) = \sum_{s,t} I_E(r-s,c-t)\, g_{m,n}(s,t) \tag{12}$$

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. We used the following parameters: 5 scales (*S*=5), and 6 orientations (*K*=6).In our model we used Gabor filter for texture segmentation. In the implementation phase, each image is divided into 4x4 blocks, and each block is convolved with Gabor filter. 6 orientations and 5 scales Gabor filters are used, and the mean and variance of the Gabor sizes are calculated for each block. In our project, Using Gabor filter, we can identify five different textures and their identities as soft, rough, smooth, porous, wavy.

B. 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. Afterwards, the food item has to be classified, using SVM [8][9], which is one of the popular techniques used for data classification. A classification task usually involves training and testing data, which consist of some data instances. Each instance in the training set contains one class label and several features. The goal of SVM is to produce a model that predicts target value of data instances in the testing set, which are given only by their attributes. In our model, we use the radial basis function (RBF) kernel, which maps samples into a higher dimensional space in a non-linear manner. Unlike the linear kernels, the RBF kernel is well suited for the cases in which the relation between class labels and attributes is nonlinear. In our proposed method, the feature vectors of SVM contain 5 texture features, 5 color features, 3 shape features, and 5 size features. The feature vectors of each food item, extracted during the segmentation phase, will be used as the training vectors of SVM. All details are explained in [10].

IV. EXPERIMENTAL RESULTS

We have applied color and texture segmentation in [2]. We applied our method to 3 different categories of food: single food, non-mixed food, and mixed food.

In [2], we achieved better accuracy for non-mixed food, and our method had problems in detecting some mixed foods. As an example, illumination of food portions in a mixed food may change as they get mixed, making it harder to extract different food portions. Furthermore, the size of food portions in different mixed food are not similar, hence the method fails to segment food portions properly. To solve these problems, in this paper we have applied graph cut segmentation to improve our segmentation step. We also trained the system with more mixed food to get more accurate results.

In this paper, our data set comprises of 15 different categories of food and fruits. These food and fruit images are divided into training and testing set, where around 50% of the fruit images from each group are used to train the system and the remaining images serves as the testing set.

As the Figure 4 shows, by applying graph cut segmentation, we have better region boundaries in our segmentation phase, which helps us to extract more accurate features from segmentation phase. The results of this step feed into our SVM model to recognize each food portions. As we can see in Table I, we have an approximately 3% increase in our single food classification accuracy in comparison to our previous model [10]. In addition to single food, we applied our method to 2 other categories of food: non-mixed food (e.g. steak and fries), and mixed food (e.g. curry or soup). The resulting SVM's accuracy is approximately 90% and 75%, respectively. In comparison to our previous model, we have achieved 5% and 15% increase in accuracy for non-mixed and mixed food, respectively.

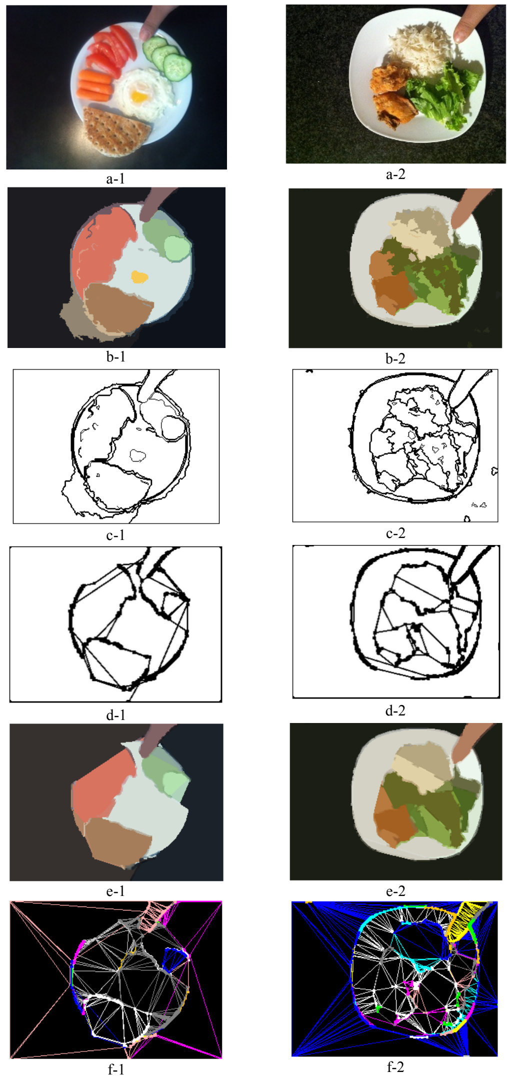

Figure 4.a-1,a-2: Original Image, b-1,b-2: Image reconstruction from segmentation, c-1,c-2: Segmentation boundaries, d-1,d-2: Image of the graph, e-1,e-2: Image reconstruction from the graph, f-1,f-2: Region based triangulation of the graph.

V. CONCLUSION

In this paper, we have proposed a method for extracting graphs and texture from food images using Graph cut segmentation. We performed a number of experiments where we fed the results from the segmentation module to the classification module. Our experiments show an overall accuracy increase of 3% in 30 different single food portions, 5% in non-mixed food, and 15% in mixed food, compared to our previous work [10]. In our future work, we are going to increase our data base and focus more on mixed food classification.

Table I. Food Recognition Accuracy for Single Food

No.	Food items	Recognition Rate (%)	
		Using color-texture segmentation	Using graph-cut, color-texture segmentation
1	Red Apple	97.64	100
2	Orange	95.59	97.5
3	Corn	94.85	96
4	Tomato	89.56	95
5	Carrot	99.79	100
6	Bread	98.39	99
7	Pasta	94.75	98
8	Sauce	88.78	92
9	Chicken	86.55	89
10	Egg	77.53	83
11	Cheese	97.47	97
12	Meat	95.73	96
13	Onion	89.99	93
14	Beans	98.68	98
15	Fish	77.7	85
16	Banana	97.65	97
17	Green Apple	97.99	97
18	Cucumber	97.65	98
19	Lettuce	77.55	85
20	Grapes	95.7	95
21	Potato	88.56	89
22	Tangerine	97.59	99
23	Chocolate Cake	88.19	85
24	Caramel Cake	85.29	85
25	Rice	94.85	94
26	Green Pepper	97.99	98
27	Strawberry	83.47	98
28	Cooked Vegetable	92.62	96
29	Cabbage	77.55	100
30	Blueberry	83.47	95
	Total average	**92.21**	**95**

VI. REFERENCES

[1] W. H. Organization, "Obesity Fact Sheet," [Online]. http://health.india.com/news.

[2] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, "An Image Processing Approach for Calorie Intake Measurement," *in Proc. IEEE Symposium on Medical Measurement and Applications*, pp. 1-5, 2012.

[3] P. Pouladzadeh, S. Shirmohammadi, A. Bakirov, and A. Bulut, "Cloud-Based SVM for Food Categorization", *Multimedia Tools and Applications, Springer,* (accepted, to appear).

[4] A. k. jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 24, pp. 1167-1186, 1991.

[5] Y. B. Yuri, G. F. Lea, "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision*, vol.70, no.2, pp.109-131, 2006.

[6] Y. Boykov, V. Kolmogorov, "An experimental comparison of mincut/max-flow algorrithms for energy minimization in vision," *IEEE transaction PAMI*, vol. 26, no. 9, pp. 1124-1137, 2004.

[7] P. Das, O. Veksler, "Semiautomatic Segmentation with Compact Shape Prior," *Journal on Image and Vision Computing,* vol 27, no.2, pp. 206–219, 2009.

[8] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge*, vol. 2, no. 2, pp. 121–167, 1998.

[9] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.

[10] P. Pouladzadeh, S. Shirmohammadi, and R. Almaghrabi, "Measuring Calorie and Nutrition from Food Image", *IEEE Transactions on Instrumentation and Measurement,* 10 pages, 2014.

[11] J. Pollak, G. Gay, S. Byrne, E. Wagner, D. Retelny, L. Humphreys, "It's Time to Eat! Using Mobile Games to Promote Healthy Eating," *IEEE Pervasive Computing*, vol.9, no.3, pp. 21-27, 2010.

[12] T. Miyazaki, G.C. De Silva, K. Aizawa, "Image-based Calorie Content Estimation for Dietary Assessment," *IEEE International Symposium on Multimedia (ISM)*, pp. 363-368, 2011.

[13] H. C. n Chen, W. Jia; Z. Li; Y. Sun, M. Sun, "3D/2D model-to-image registration for quantitative dietary assessment," *Northeast Bioengineering Conference (NEBEC)*, pp. 95-96, 2012.

[14] M. Sun, Q. Liu, K. Schmidt, J. Yang, N. Yao, J. D. Fernstrom, M. H. Fernstrom, J. P. DeLany and R. J. Sclabassi, "Determination of Food Portion Size by Image Processing," *International IEEE EMBS Conference*, pp. 871 - 874, 2008.

[15] Y. Saeki and F. Takeda, "Proposal of Food Intake Measuring System in Medical Use and Its Discussion of Practical Capability," *Springer-Verlag Berlin Heidelberg*, vol. 3683, pp. 1266–1273, 2005.

[16] C. K. Martin, S. Kaya and B. K. Gunturk, "Quantification of food intake using food image analysis," *International Conference of IEEE Engineering in Medicine and Biology Society*, pp. 6869-6872, 2009.