Segmentation and Recognition of Multi-Food Meal Images for Carbohydrate Counting

Marios Anthimopoulos *Member, IEEE,* Joachim Dehais, Peter Diem and Stavroula Mougiakakou *Member, IEEE*

***Abstract*— In this paper, we propose novel methodologies for the automatic segmentation and recognition of multi-food images. The proposed methods implement the first modules of a carbohydrate counting and insulin advisory system for type 1 diabetic patients. Initially the plate is segmented using pyramidal mean-shift filtering and a region growing algorithm. Then each of the resulted segments is described by both color and texture features and classified by a support vector machine into one of six different major food classes. Finally, a modified version of the Huang and Dom evaluation index was proposed, addressing the particular needs of the food segmentation problem. The experimental results prove the effectiveness of the proposed method achieving a segmentation accuracy of 88.5% and recognition rate equal to 87%.**

I. Introduction

THE global spread of metabolic disorders such as obesity and diabetes has raised strong concerns and an urgent need for dietary intake monitoring and control systems. Conventional mobile applications for dietary advice usually involve a great amount of human interaction while they often introduce significant errors due to the inability of patients to assess accurately their food intake. Recently, the widespread use of smartphones with enhanced capabilities together with the recent advances in image analysis enabled the development of a new generation of automatic dietary assessment systems based on computer vision techniques.

A typical computer vision based, dietary assessment system consists of four basic sub-modules: food segmentation, food recognition, volume estimation and nutritional analysis. In the current study, we propose two novel methods for the segmentation and recognition of multiple-food images as the first steps of a system providing automatic carbohydrate counting and prandial insulin bolus advice to Type 1 diabetic (T1D) patients (Fig. 1).

Manuscript received July 30, 2013. This work was supported in part by the Bern University Hospital, "Inselspital" and by the European Union Seventh Framework Programme (FP7-PEOPLE-2011-IAPP) under grant agreement n° 286408 [www.gocarb.eu].

M. Anthimopoulos and J. Dehais are with the ARTORG Center for Biomedical Engineering Research, University of Bern, 3010 Bern, Switzerland (emails: {marios.anthimopoulos, joachim.dehais} @artorg.unibe.ch) .

P. Diem is with the Bern University Hospital, Inselspital, Dep. of Endocrinology, Diabetes and Clinical Nutrition, 3010 Bern, Switzerland (email: peter.diem@insel.ch).

S. G. Mougiakakou is with the University of Bern, 3010 Bern, Switzerland (Corresponding author; phone: +41 31 632 7592; fax: +41 31 632 7576; email: stavroula.mougiakakou@artorg.unibe.ch).

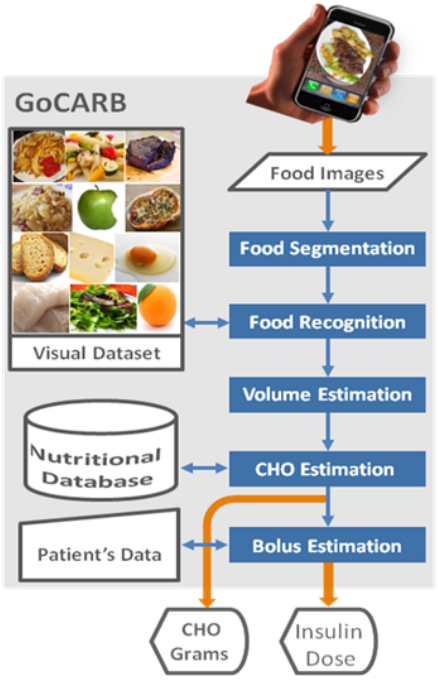


Fig. 1. Architecture of the entire system aiming at carbohydrate counting for T1D patients.

Prandial is the insulin dose used in order to cover the effect of a meal and is closely related to the carbohydrate (CHO) intake. Clinical studies have shown that in children and adolescents on intensive insulin therapy a ±20 g variation significantly impacts the postprandial glycaemia [1] leading potentially to severe long-term complications [2]. Furthermore, according to recent studies, a substantial number of adult T1D patients estimate prandial insulin needs inappropriately [3-4].

II. Related Works

Several dietary assessment systems have been proposed during the last 5 years including food segmentation and/or recognition modules, mainly for the prevention and treatment of obesity.

For the segmentation of easily separated food items in a white plate, adaptive thresholding [5] and connected component analysis [6] have been used. In [7], food items were segmented by matching the Scale Invariant Feature Transform (SIFT) points of the image with a reference image dataset. Furthermore, densely multi-class classification was proposed for segmentation via the generation of saliency

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

maps [8]. In [9] food images were segmented using normalized cuts based on intensity and color. Finally, in [10] the detection of several candidate regions was proposed by fusing outputs of several region detectors including Felzenszwalb's deformable part model (DPM) [11], a circle detector and the JSEG region segmentation [12].

Various feature sets and classification techniques have been used for the recognition of the already segmented food items. In [5], food items were described by simple color, size, texture, shape and context-based features. Average color values and Gabor features were used in [6] while in [7], food items were described by a bag of SIFT features. Furthermore, color neighborhood and maximum response features have been used within a bag-of-features model [8]. Bag-of SIFT and CSIFT features with spatial pyramid, histogram of oriented gradient, and Gabor filter responses were used in [10]. For the classification, Support Vector Machines (SVM) [6], [8], [10] Artificial Neural Networks [5] and Bayesian classifiers [7] are among the popular choices.

In this work, we propose the use of mean-shift algorithm [13] for the food segmentation. Mean-shift has proved its superior performance in various segmentation applications having though a relatively high computational cost. To this end, we adopt a pyramidal approach which reduces substantially the computational cost and together with the proposed LAB-based color distance achieves both accuracy and efficiency. An index for the food segmentation evaluation is also proposed. For the recognition, the complementary information from learned color histograms and Local Binary Patterns provide an informative description that lead to high recognition rates.

III. Proposed Methods

The proposed system consists of two distinct modules for segmentation and recognition. The input of the system is a picture of a served meal on a circular plate (Fig. 2a) whereas the output is a map with specific label codes for the background and the different food classes (Fig. 2b).

A. Food Segmentation

The food segmentation algorithm is based on color information and consists of five main steps: CIELAB conversion, pyramidal mean-shift filtering, region growing, region merging, and plate detection/background subtraction.

(a) (b)

Fig. 2. (a)The system's input and (b) the desirable output.

1) CIELAB conversion

Initially, the input image is converted to the perceptually uniform CIELAB color space. The Euclidean distance between two colors in CIELAB is considered more representative of their visual color difference. Moreover, the separate lightness channel (L) provides useful properties as explained later.

2) Pyramidal mean-shift filtering

Mean-shift is an iterative algorithm for feature space cluster analysis which has been applied with great success to image segmentation [13]. For the image segmentation problem we consider as feature space, the joint space-color hyperspace of five dimensions consisting of the two spatial coordinates (X, Y) and the three color channels (L, A, B).

The algorithm consists of three iterating steps:

1. At every pixel $P=[(X, Y), (L, A, B)]$ of the image, a neighborhood is defined in the hyperspace by the points:

$$[(x,y),(l,a,b)]:\begin{cases}|X-x|\le sth,\\ |Y-y|\le sth,\\ \|(L,A,B),(l,a,b)\|\le cth\end{cases} \quad (1)$$

where $$\|(L,A,B),(l,a,b)\|=\sqrt{|L-l|+(A-a)^2+(B-b)^2} \quad (2)$$

and sth, cth are parameters experimentally estimated.

2. Mean vector P' is calculated over the neighborhood.

3. P' is used as center point of the next iteration.

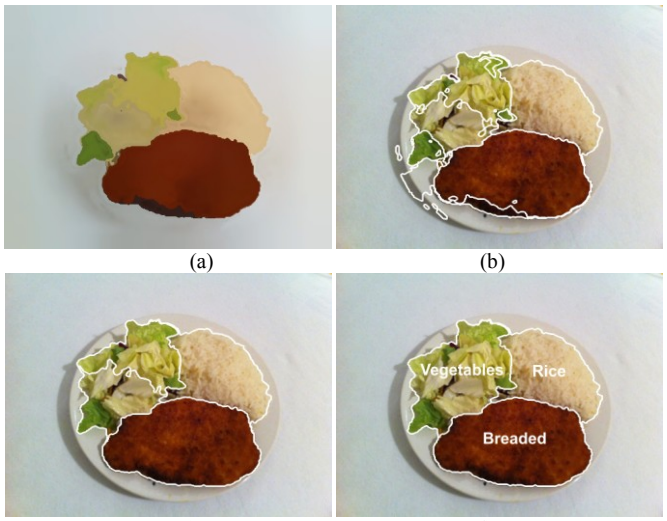
The algorithm iterates until a certain number of iterations is reached or the shift of the center point becomes very low. After finishing the iterations, the starting pixel gets the final mean color. The distance of (2) was chosen instead of the Euclidean in order to reduce the significance of lightness and eliminate the shadow effect.

In order to enhance the method's efficiency, a Gaussian pyramid is constructed with four levels, and the algorithm is applied on the smallest scale first. After that, the results are propagated to the larger scale and the iterations are run again only on pixels with a color distance more than *cth* from at least one neighbor.

After filtering, the fine-grain texture is smoothed without losing though the dominant color edges (Fig. 3a). Hence, pixels of the same segment ideally have similar colors and distinguishable from the rest of the segments. If the previous assumption is true then a region growing algorithm could grow any seed pixel to the entire area of the corresponding segment.

3) Region Growing

The proposed region growing algorithm chooses seeds randomly from the pixels which have not been assigned yet to any segments, and expands them to all directions when the color distance (2) of a neighboring pixel is less than 0.6**cth* from the average segment color. After region growing the initial segmentation is produced (Fig. 3b). However, many of the produced segments are too small and assuming there is a minimum size of food item we can proceed with a region



(a) (b) (c) (d)

Fig. 3. The processing stages of the proposed system for the image of Figure 2a. (a) mean-shift filtering result, (b) region growing result, (c) region merging result, (d) recognition result.

merging step.

4) Region Merging

In this step, every region with an area lower than a threshold *ath* is merged with the adjacent segment with the minimum color distance (2). The final segmentation result is presented in Fig. 3c. After merging the small regions, some food items may remain over-segmented. However, they will be automatically merged after recognition (Fig. 3d).

5) Plate Detection/Background subtraction

Before proceeding to the recognition stage the system should specify which of the produced segments correspond to food items by discarding the background segments. To this end, we locate the plate in the image by using an ellipse detector. An edge map is created, edge components with less than 16 pixels are discarded and then the RANSAC paradigm [14] is applied given the ellipse-generating property of single components. Each segment with more than 10% of its area outside the ellipse is considered background. Furthermore, each of the remaining segments that shares borders with the background for more than 10% of its contour's length is labeled as plate region and it is also discarded.

B. Food Recognition

The recognition module consists of two stages: description and classification. In this study we use color and texture feature sets for the food description and both sets are histogram-based so they can be easily computed regardless the segment's shape, and normalized based on its size.

As for color features, the histogram of the 1024 most dominant food colors was used. To this end, a hierarchical version of the k-means [15] algorithm is applied to cluster the color space created by the training set of food images. Thus, the centers of the generated clusters represent the considered dominant colors. The hierarchical k-means was chosen instead of the original k-means for the sake of efficiency during the calculation of features.

For texture features the Local Binary Pattern (LBP) operator was used [16]. The LBP operator consists of a 3x3 kernel where the centre pixel is used as a threshold. Then the eight binarized neighbors are multiplied by the respective binomial weight producing an integer in the range [0-255]. Each of the 256 different 8-bit integers is considered to represent a unique texture pattern. Thus, the LBP histogram values of an image region describe its texture structure.

After combining color and texture features, a vector of 1280 dimensions is created and fed to a non-linear SVM with a Radial Basis Function (RBF) kernel that will assign to the segment one of six predefined food classes.

IV. Experimental Results

The performance of the two proposed modules was evaluated separately since testing the food segmentation needs images with multiple foods served on a plate whereas recognition requires a huge amount of images without any restriction on the food serving.

A. Segmentation

The segmentation dataset consists of 65 images captured with various cameras by the authors and annotated manually for the creation of a ground truth map (Fig. 2b). For the evaluation of the food segmentation algorithm, we propose a metric similar to the Huang and Dom index (HDI) [17]. Let $S = \{S_1, S_2, \ldots, S_m\}$ and $T = \{T_1, T_2, \ldots, T_n\}$ be two segmentations, where S_i/T_i is the set of pixels in region i from segmentation S/T and m, n are the number of segments in S and T, respectively. The normalized directional index is defined as:

$$NI_H(T \Rightarrow S) = \frac{\sum_{S_i \in S} Max_j\left(\left|S_i \cap T_j\right|\right)}{\sum_{S_i \in S} \left|S_i\right|} \quad (3)$$

HDI is defined as the arithmetic mean of the two reverse normalized indices and although it has been reported as the most objective measure for supervised evaluation of image segmentation [18] it still presents one drawback. The use of the simple arithmetic mean cannot capture the imbalance between the two reverse normalized indices, which is an indication of a bad segmentation result. To this end, we propose the use of the harmonic mean instead of the arithmetic:

$$HSI = \frac{2 * NI_H(T \Rightarrow S) * NI_H(S \Rightarrow T)}{NI_H(T \Rightarrow S) + NI_H(S \Rightarrow T)} \quad (4)$$

Finally, we propose the exclusion of background segments from the computation of the indices (3). Hence, the result becomes independent from the size of the background area while the performance of the background detection method is incorporated in the evaluation procedure.

Table I presents the overall results of the different

segmentation variants together with the average computational time per image. The results prove the ability of the LAB-based color distance to deal with irregular lighting conditions and justify the choice of the pyramidal filtering for reducing the computational cost while further improving the system's accuracy. Based on HSI (4), the optimal system thresholds were estimated: *cth* = 11, *sth* = *plate_radius*/5, *ath* = *image_area*/50. The proposed segmentation method incorporates parts of the OpenCV source code [19].

TABLE I
SEGMENTATION ACCURACY RESULTS

Method	*HSI*	Secs/image
RGB/No-pyramid/Euclid. color dist	0.705	11
LAB/No-pyramid/ Euclid. color dist	0.769	8
LAB/No-pyramid/Proposed dist (2)	0.875	13
LAB/Pyramid/Proposed dist (2)	**0.885**	**2.8**

B. Recognition

For training and testing the food recognition stage more than 5000 food images were gathered from the web and manually annotated. The dataset's annotation allowed the creation of a new dataset of over 13000 image patches belonging to the six major classes of international food: meat, breaded food, rice, pasta, potatoes and vegetables. Using the latter dataset and a 10-fold cross-validation model, the system's accuracy was estimated in the order of 87% and the corresponding confusion matrix was computed (Fig. 4). The results prove the effectiveness of the proposed method in recognizing food types with relatively distinct color and texture like breaded food, vegetables and meat. On the other hand, distinguishing between pasta, potatoes and rice seems more challenging mainly due to color similarity between the first two and the use of similar dressings for pasta and rice.

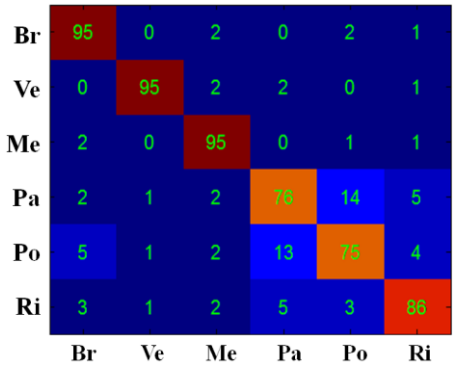


Fig. 4. Confusion matrix of the proposed classification algorithm for the classification of six food classes (**Br** = breaded food, **Ve** = vegetables, **Me** = meat, **Pa** = pasta, **Po** = potatoes, **Ri** = rice).

V. CONCLUSION

In this study, we propose a system for the automatic segmentation and recognition of multi-food images towards CHO counting for T1D patients. The provided results prove the effectiveness of the methods despite the challenges of the problem. As future work, we plan to increase the number of food classes and enhance the segmentation algorithm by exploiting additional texture features and depth information.

REFERENCES

[1] C. E. Smart, B. R. King, P. McElduff, and C. E. Collins, In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia, *Diabetic Medicine*, 2012.

[2] R. Peter, OE. Okoseime, A Rees, DR Owens, Postprandial glucose - a potential therapeutic target to reduce cardiovascular mortality, *Current Vascular Pharmacology*, vol. 7 no.1, pp. 68-74, 2009.

[3] A. J. Ahola, S. Mäkimattila, M. Saraheimo, V. Mikkilä, C. Forsblom, R. Freese, P.-H. Groop, Many patients with type 1 diabetes estimate their prandial insulin need inappropriately, *Journal of Diabetes*, vol. 2, no. 3, pp. 194-202, 2010.

[4] F. K. Bishop, D. M. Maahs, G. Spiegel, D. Owen, G. J. Klingensmith, A. Bortsov, J. Thomas, and E. J. Mayer-Davis, The carbohydrate counting in adolescents with type 1 diabetes (CCAT) study, *Diabetes Spectrum*, vol. 22, no. 1, pp. 56–62, 2009.

[5] G. Shroff, A. Smailagic, and D. P. Siewiorek, Wearable context-aware food recognition for calorie monitoring, in *12th IEEE International Symposium on Wearable Computers*, pp. 119–120, 2008.

[6] F. Zhu, M. Bosch, I. Woo, S. Y. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp, The use of mobile devices in aiding dietary assessment and evaluation, *IEEE Journal of selected Topics in Signal Processing*, vol. 4, no. 4, pp. 756–766, 2010.

[7] F. Kong and J. Tan, DietCam: Automatic dietary assessment with mobile camera phones, *Pervasive and Mobile Computing*, pp. 147–163, Feb. 2012.

[8] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, Recognition and volume estimation of food intake using a mobile device, in *Workshop on Applications of Computer Vision (WACV)*, pp. 1–8, 2009.

[9] F. Zhu, M. Bosch, T.E. Schap, N. Khanna, D.S. Ebert, C.J. Boushey, and E.J. Delp, Segmentation assisted food Classification for dietary assessment, *Proceedings of the IS&T/SPIE Conference on Computational Imaging IX*, Vol. 7873, San Francisco Airport, California, January 2011.

[10] Y. Matsuda, H. Hoashi, K. Yanai, Recognition of Multiple-Food Images by Detecting Candidate Regions, *Multimedia and Expo (ICME), 2012 IEEE International Conference on* , vol., no., pp.25-30, 9-13 July 2012.

[11] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[12] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of colortexture regions in images and video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 800–810, 2001.

[13] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.24, no.5, pp.603-619, May 2002 doi: 10.1109/34.1000236.

[14] M. Fischler, R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". *Commun. ACM* 24, 6 ,381-395, 1981

[15] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[16] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[17] Q. Huang and B. Dom, Quantitative methods of evaluating image segmentation, *International Conference on Image Processing (ICIP)*, pp 53–56, 1995.

[18] K. Mc Guinness, E. O'Connor, Image segmentation, evaluation, and applications. PhD Thesis, Dublin City University [Online]. 2010. Available from: http://doras.dcu.ie/14998/

[19] The OpenCV Library [Online]. Available: http://opencv.org/