Journal of Engineering and Applied Sciences 12 (8): 2200-2206, 2017
ISSN: 1816-949X

Aggregated Features Association Classifier for Multiple Food Items Identification

[1, 2]Salwa Khalid Abdulateef, [1]Massudi Mahmuddin and [1]Nor Hazlyna Harun
[1]School of Computing, College of Art and Sciences, Universiti Utara Malaysia, Changlun, Malaysia
[2]Department of Computer Science, College of Computer and Mathematics Sciences,
Tikrit University, Tikrit, Iraq

Abstract: Image based food identification is an emerging research topic for much industrial application. It refers to the capability of identifying various food items based on the visual information. Unfortunately, food items classification is highly sensitive to the accuracy of the image segmentation which is not always satisfying due to many factors. In this study, an aggregated features association classifier is proposed to handle the resultant problem of non-accurate image segmentation. It uses ELM for food items classification. Also, it exploits the fact that food items are associated with others when they are placed in the plate; the accuracy of the classifier has been improved using features association. An accuracy of 100% is obtained for input images with over or under segmentation errors which proves the usefulness of this algorithm.

Key words: Image food identification, extreme learning machine, feature extraction, object recognition, image analysis

INTRODUCTION

Food identification is becoming an interesting research topic now a days (Khanna *et al*., 2010; Chen *et al*., 2012). The aim is to facilitate understanding the ingredient of the dietary preferences and o supply medical experts with reliable measure of their patient's food intake (Wu and Yang, 2009). Food identification is an essential core in any calories estimation or counting system. In order to do food identification in an accurate way, accurate image segmentation is needed. The segmentation process is required to partition a food image into different areas in order to distinguish the object in interest which represents specific food item from background and separating each object from others (Luccheseyz and Mitray, 2001; Vartak and Mankar, 2013). The main purpose of the segmentation is to ensure that only the object of interest will be processed during the object analysis or identification phase. A false segmentation will cause degradation of the estimation of calories and classification process (Pouladzadeh *et al*., 2013). Unfortunately, image segmentation is not always satisfying due to several obstacles some of them are related to variability of food items in terms of shape and arrangement while others are caused because of the overlap of food items while they are positioned in the plate. As a result, inaccurate segmentation of food items leads to failure in the identification process and subsequently, this might jeopardize the whole calories estimation system (He *et al*., 2013).

In this study, we propose aggregated feature association algorithm of food items for the sake of handling corrupted or non-accurate image segmentation at the level of identification of food items. It is assumed that some objects in the plate are non-accurately segmented. Scenarios of over-segmentation and under segmentation have occurred. When this non-accurate segmentation result is fed into the classification layer, more possibilities of misclassifications are subject to obtain. In our aggregated feature association classification algorithm, it is expected that those non relevant segmentation errors of over-segmentation or under-segmentation of food items can be handled. To best of our knowledge, this work is the first object identification work that assumes or accepts non-accurate input of image segmentation and tries to overcome these errors by using a novel features association algorithm.

Various classification approaches of food items in image have been developed in the literature. In computer vision, food recognition is a specific case of category recognition. Martinel *et al*. (2015) have developed a

Corresponding Author: Salwa Khalid Abdulateef, School of Computing, College of Art and Sciences, Universiti Utara Malaysia, Changlun, Malaysia

multi-agent based food recognition system based on optimal features extracted from large set of features (e.g., color, texture). Each agent is an extreme machine that is trained on one feature type for identifying one food plate. Experiments prove that this system outperformed other state-of-the-art works of public benchmark datasets. Discriminative visual features such as color, shape and texture are computed by the feature extraction model.

Joutou and Yanai (2009) and Hoashi *et al.* (2010) a system for identifying menu food images has been developed. Next, manual clipping is used to segment the objects of food. Classifier is trained on set of extracted features are color, texture, Scale Invariant Feature Transformation (SIFT) and gradient. The accuracy turned out to be highly dependent on food categories because of similarity of some categories in one food image. Support Vector Machine (SVM) was the most used type of classifier among all different types of classifiers in food images. The researchers have applied the first system to carbohydrate counting and insulin advisory for diabetic patients by using automatic segmentation and recognition image for the segment food image used pyramidal mean-shift filtering, a region growing algorithm and region merging then recognition step initially extracted color and texture features then entered it to SVM with a Radial Basis Function (RBF) kernel for classification. The system scored 88.5% for segmentation accuracy and recognition ratio was equal to 87% and time was 2.8 sec/image but have limitation specified only carbohydrate food type and must plate, background the same color (Anthimopoulos *et al*., 2013). Anthimopoulos *et al.* (2014) an optimized system for food recognition has been developed. The goal is to estimate a meal's carbohydrate content for diabetic patients rely on using the scale-invariant feature transform on the HSV color space, builds a visual dictionary of 10000 visual words by using the hierarchical k-means clustering then used support vector machine to classifier each type of food, the results showed the SIFT perform significantly better than color descriptors, probably because they are less sensitive to intensity changes and color shifts. The system score the accuracy is 78% but misclassified four categories of food as eggs, cheese, pizza and pasta.

The research by Zhu *et al.* (2015) has presented a quantitative evaluation for segmentation and classification also applied the system of dietary assessment by different method to analysis image which is multiple hypotheses of image segmentation to select stable segmentations based on the classifier's confidence score for each region segmented. The idea was partitioned the objects into perceptually similar object classes based on global and local features; to assess the accuracy of image segmentation. K-Nearest Neighbors (KNN) support vector machines techniques used to classify the feature channels, the drawbacks are the iteration number of segment depend on size of image, a specific dataset used, depend on color checkerboard pattern is placed in the scene to ensure color consistency due to changes in illumination and viewpoints and this method not compared.

The researchers have been extracted color, texture, LBP and SIFT to identified50 China 3D food images which captured by camera kinect, then classified by multi class SVM (non linear) is trained for each feature, the accuracy is 68% (Chen *et al*., 2012). Matsuda *et al.* (2012) have developed food recognition system for multiple food items in one image in the same time. The method consists of sliding window search, detect food by circular plate, region segmentation by JSEG, then extracted color histogram, SIFT, spatial-pyramid, texture and shape features then feed them into Multiple Kernel Learning (MKL) and support vector machine to classification. The experiments on 1200 Japanese food item and achieved 55.8% classification rate and processing time 2 sec, however this method has constraint as plate must circular, Japanese food, sliding window search rely on gradient based feature only, segment not good if region = ten making overlap between background and food item and combined region based on circularity.

The research by Kong and Tan (2012) has presented a system for fast food (homemade food and fruits) recognition based on camera dietcam mobile phone to Food intake assessment by take only three picture or record a short video. To recognition used SIFT features clustered into visual words and put in Bayesian probabilistic for classification, although the performance of recognition is 92% but the system is limited for types of food only two and the number of food items to be recognized is less than six. It is important to notice that the accuracy of the classification that was reported in previous works is not reliable as it only indicates to compare between the classifiers due to different factors. Some of them are related to the simplicity of the scenarios that is conducted for validation in some approaches and because of providing different assumptions for the experiment. But most importantly, some of the methods were not really food items classification as food categorizations such as developed the web application system named foodlog (Maruyama *et al*., 2010). Shroff *et al.* (2008) the features shape and texture are extracted and provided to two layers feed forward neural network NN. Back propagation learning algorithm has been used for training NN. Four outputs provided to NN for classification of four food classes. The concept of contextual information usefulness in improving the accuracy of classification was validated in this research. Two types of contextual information are used (context profile and session context). This research aims atusing the data logging of the user in the smartphone as contextual information provider to overcome challenges and complexity in food classification. The platform that is developed for this purpose is DiaWear. However, the system is still primary in the aspect of image processing, and segmentation part. In addition, it is very limited in terms of food types.

In the training mode, the goal is to find out the appropriate values of the weights based on given training data set that is considered to have enough knowledge about the application nature. Once the training mode is finished, the ANN uses its final value of weights to provide an output according to given input (Basu *et al*., 2010). There are many determinates involve in the performance of the ANN, the structure, the number of inputs and ANN training. Unfortunately, ANN training sometimes is costly and time consuming, other negative aspects of ANN is over-fitting (Savakar, 2012). Yang *et al*. (2010) have been taken 61 categories different ingredients based food classification method that achieved 78% accuracy. Food ingredients were detected through texture classification and food types were classified by calculating the pairwise statistics between food ingredients.

A classification of food images by incorporating temporal information and employ recursive Bayesian estimation to incrementally learn from a person's eating history has been developed by Wang *et al*. (2015). Results show an improvement of food classification accuracy by 11%. Multi-stage system has been developed by Pouladzadeh *et al*. (2014). First stage is scaling size of the food images; second one is applying k-means iteratively to calculate the set of clusters. Next, features of color segment and texture segments are extracted. SVM classifier has been used and estimation of food calories is done based on the area of the amount of each item. Total accuracy of 92% for simple type of food only is performed. From the presented literature, it can be observed that all the previous works have built classification algorithm based on classifier trained on individual items features or mixture of items in one plate. However, no consideration has been taken to the habit or culture aspect in association between multiple food items as assisting factors in food item classification.

MATERIALS AND METHODS

This study presents the methodology developed for aggregated feature association classification of multi-food items. In the next study an overview of ELM is provided. Next, the block diagram of the system as a whole is presented. Finally, the section of feature extraction is introduced.

Overview of Extreme Learning Machine (ELM): In order to classify the items, ELM has been used. ELM is a kind of feed-forward single hidden layer neural network whose input weights and thresholds of hidden layers are generated randomly. Because the output-weights of ELM are calculated by the least-square method, the ELM presents a high speed on training and testing. Considering that each vector of features is regarded as one row in the data and its corresponding item encoding. The data consists of N distinct samples (x_j, t_j). Where, $j = 1, ..., N$, $x_j = (x_{j1}, x_{ij}, ..., x_{jn})$. Standard Single Hidden Layer Feed-Forward Network (SLFN) can be modeled with an activation function g (x) and $\tilde{N}$ hidden layer neurons as following Eq. 1:

$$\sum_{i=1}^{\tilde{N}} \beta_i \, g(w_i \, x_j + b_i) = t_i$$

$$j = 1, ..., N, \; w_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{in})^T$$

Where:

b_i = The threshold of the hidden node
β_i = The weight connecting the ith hidden node and the output

It has been proved by Huang *et al*. (2006) that if the activation function is differentiable then the required number of the hidden layer neurons is lower than the data size or $\tilde{N}$<N.

The training algorithm of the ANN described above is introduced in three steps as following:

- Assign randomly random weights and biases as w_i and b_i
- Calculate the hidden layer output matrix
- Calculate the output weights by using the Moore-Penrose generalized inverse of hidden layer output matrix

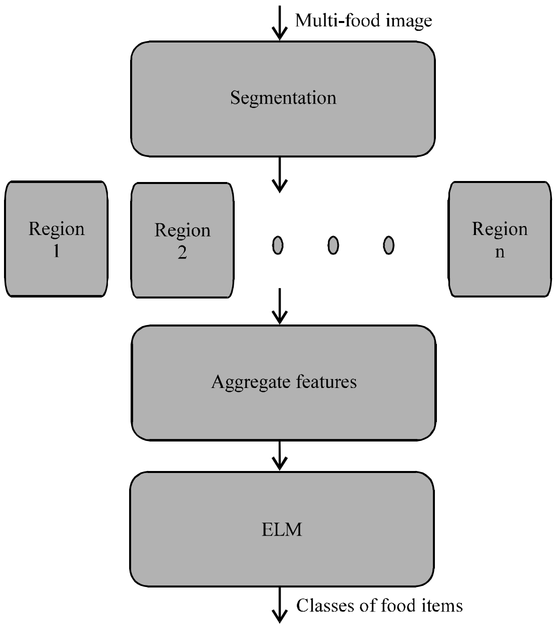

Fig. 1: Block diagram of the system

Other advantages of this approach are its simplicity, efficiency in terms of execution because it has one training iteration, more capability of capturing knowledge from training dataset. Moreover, this approach behaves well with many non-linear activation and kernel functions (Rajesh and Prakash, 2011; Huang, 2014). Researchers have considered this approach simpler than other for feed forward neural networks (Huang *et al*., 2006). One of the good features is the independency in the hidden nodes parameters from the training data set. While the output nodes are calculated analytically. This is unlike traditional ANN based learning algorithms. Another appealing advantage of this approach is having two modes of functionality: regression and classification.

Block diagram of aggregated features of multi food classification system: The block diagram of the system is shown in Fig. 1. As it is observed, the features are extracted from the whole items from left to right and aggregated in one bag of features to be fed into ELM classifier. This is useful to add to the classifier an awareness of the associated items and food arrangement for each item. In case the segmented image is not accurate because of over segmentation or under-segmentation, the classifier will have more capability to identify food items relying on the knowledge provided by other features associated with the subject food item.

Feature extraction: Each region represents a food item. In order to identify the item, features must be extracted. The aim was to combine comprehensive knowledge about food item color, shape, size and texture. These features are combined in one vector called bag of features and augmented with other bags of features for all regions in the plate.

Total number of features has been extracted are 32 features as following: color features (11 features): mean value of the following color components: I1, I2, I3, normalized R, normalized G, normalized B, H from HSV, Cb, Cr, a*, b*. Contours fitting features: the contour of an item is split into four segments after centered on zero. Each segment is fit with a curve from the third degree. Each curve has four coefficients. Therefore, these features together are 12 features. Geometrical features: there are five features as:

- Size of item
- Radius of the circle that has the same area with item
- Major and minor axis of the smallest ellipse containing the item
- Extent: area of the item divided by the area of bounding box

Texture features: firstly, compute 16 co-occurrence matrices using the following offsets:

offsets = 01; 02; 03; 04; ...
-11; -22; -33; -44; ...
-10; -20; -30; -40; ...
-1-1; -2 -2; -3 -3; -4 -4];

and then computing the four following features as mean value from the 16 matrices: contrast, correlation, energy and homogeneity.

RESULTS AND DISCUSSION

The experimental work has been done on 2, 3 and 4 food items. Food items are arranged in one plate as it shown in Fig. 2. Table 1-3 show the selected groups of each plate that was provided for identification to the system and the class encoding that is used. The number of inputs in ELM method is fixed according to the number of items in the plate. However, as it has been mentioned previously, the total number of features for each item is 32 so the number of features is variable and functions to number of items in each image. In order to handle this problem, three neural networks have been built and trained using ELM. The number of inputs of the first one is 64, the number of inputs of the second one is 96 and the number of inputs of the third one is 128. The following flowchart clarifies the proposed method from

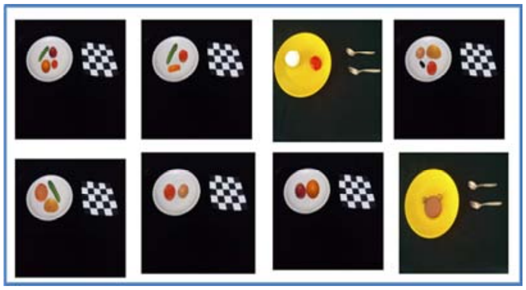

Fig. 2: Samples of dataset of two or three or four items in the image

Table 1: Food type of plate of three food items 1

Item 1	Item 2	Class
Apple	Orange	1
Cucumber	Tomato	2
Egg	Tomato	3
Banana	Apple	4
Candy	Date	5
Meat	Green olive	6
Carrot	Cucumber	7
Egg	Green olive	8
Tomato	Capsicum	9
Bagel	Cake	10

Table 2: Food type of plate of three food items 2

Item 1	Item 2	Item 3	Class
Carrot	Cucumber	Tomato	1
Bread	Meat	Tomato	2
Tomato	Black olive	Egg	3
Banana	Orange	Apple	4
Bagel	Candy	Date	5
Green olive	Tomato	Bread	6
Bread	Meat	Cucumber	7
Egg	Green olive	Black olive	8
Tomato	Carrot	Capsicum	9
Bagel	Candy	Cake	10

Table 3: Food type of plate of three food items 3

Item 1	Item 2	Item 3	Item 4	Class
Cucumber	Tomato	Apple	Orange	1
Capsicum	Carrot	Cucumber	Tomato	2
Cucumber	Bread	Tomato	Egg	3
Green olive	Bread	Meat	Egg	4
Orange	Banana	Cucumber	Apple	5
Tomato	Bread	Egg	Date	6
Meat	Cucumber	Tomato	Bread	7
Capsicum	Meat	Tomato	Bread	8
Bread	Tomato	Green olive	Black olive	9
Candy	Date	Bagel	Cake	10

start to the end. In order to select the right neural network for classification the algorithm depicted in Fig. 3 has been called. In order to verify the proposed method, a dataset of 300 images has been acquired using 8 mega pixel camera where for each class 10 images have been taken. Half of the images have been used for training and the other half have used for testing. The number of hidden neurons has been chosen to be 18 neurons. The activation function of NN neurons has been sigmoid

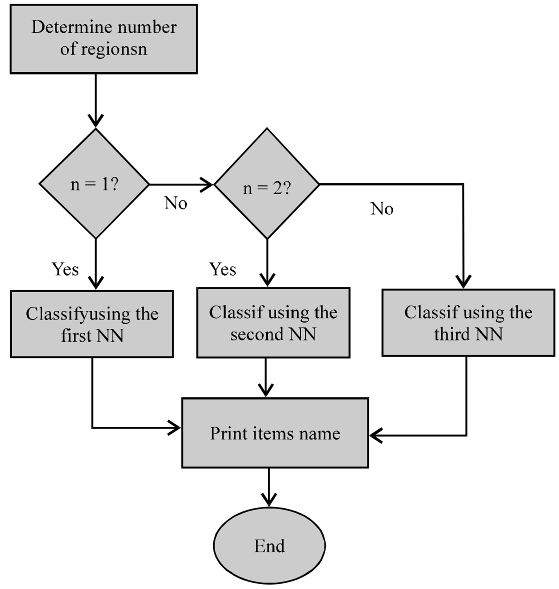

Fig. 3: Algorithm for selecting neural network for classification

Table 4: Accuracy of classification for 2-4 items

No. of items	Training accuracy (%)	Testing accuracy (%)
2	100	100
3	100	100
4	100	100

Table 5: Comparison proposed method with ELM (RBF, Sig) and SVM (linear, non linear)

Classifier type	ELM-RBF	ELM-Sig.	SVM-Linear	SVM-non linear	Aggregate ELM
Accuracy (%)	98	9	86	90	100

function. Using MATLAB environment the system has been implemented and tested. The following table shows the resultant accuracies for the three cases.

It is clear from the results that the concept of aggregated associated features classification enhances the ability of the system to classify food items which can be seen clearly from 100% accuracy for all of the case (Table 4). Comparing aggregated feature association classifier with classical ELM and SVM with different kernel in Table 5, it can be concluded that the former outperform in terms of accuracy. This is due to the aspect of incorporating associated learning in the classifier.

CONCLUSION

This research develops an aggregated features association of multiple food items identification based on extreme machine learning. It builds the classifier of the food items identification based on the knowledge of the items association. An accuracy of 100% is obtained with scenarios of 2, 3 and 4 food items in one plate. Future work is to extend the work to mixed type of food items and to implement it for wide range of food items.

REFERENCES

Anthimopoulos, M., J. Dehais, P. Diem and S. Mougiakakou, 2013. Segmentation and recognition of multi-food meal images for carbohydrate counting. Proceedings of the 2013 IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), November 10-13, 2013, IEEE, New York, USA., ISBN:978-1-4799-3164-4, pp: 1-4.

Anthimopoulos, M.M., L. Gianola, L. Scarnato, P. Diem and S.G. Mougiakakou, 2014. A food recognition system for diabetic patients based on an optimized bag-of-features model. J. Biomed. Health Inf., 18: 1261-1271.

Basu, J.K., D. Bhattacharyya and T.H. Kim, 2010. Use of artificial neural network in pattern recognition. Int. J. Software Eng. Applic., 4: 23-34.

Chen, M.Y., Y.H. Yang, C.J. Ho, S.H. Wang and S.M. Liu *et al*., 2012. Automatic Chinese food identification and quantity estimation. Proceedings of the Conference on SIGGRAPH Asia 2012 Technical Briefs, November 28-December 1, 2012, ACM, New York, USA., ISBN:978-1-4503-1915-7, pp: 1-4.

He, Y., N. Khanna, C.J. Boushey and E.J. Delp, 2013. Image segmentation for image-based dietary assessment: A comparative study. Proceedings of the 2013 International Symposium on Signals, Circuits and Systems (ISSCS), July 11-12, 2013, IEEE, New York, USA., ISBN:978-1-4799-3193-4, pp: 1-4.

Hoashi, H., T. Joutou and K. Yanai, 2010. Image recognition of 85 food categories by feature fusion. Proceedings of the IEEE International Symposium on Multimedia (ISM), December 13-15, 2010, IEEE, New York, USA., ISBN:978-1-4244-8672-4, pp: 296-301.

Huang, G.B., 2014. An insight into extreme learning machines: Random neurons, random features and kernels. Cognit. Comput., 6: 376-390.

Huang, G.B., Q.Y. Zhu and C.K. Siew, 2006. Extreme learning machine: Theory and applications. Neurocomputing, 70: 489-501.

Joutou, T. and K. Yanai, 2009. A food image recognition system with multiple kernel learning. Proceedings of the 16th IEEE International Conference of Image Processing, November 7-10, 2009, Cairo, pp: 285-288.

Khanna, N., C.J. Boushey, D. Kerr, M. Okos and D.S. Ebert *et al*., 2010. An overview of the technology assisted dietary assessment project at Purdue University. Proceedings of the 2010 IEEE International Symposium on Multimedia (ISM), December 13-15, 2010, IEEE, New York, USA., ISBN:978-1-4244-8672-4, pp: 290-295.

Kong, F. and J. Tan, 2012. DietCam: Automatic dietary assessment with mobile camera phones. Pervasive Mobile Comput., 8: 147-163.

Luccheseyz, L. and S.K. Mitray, 2001. Color image segmentation: A state-of-the-art survey. Proc. Indian Nat. Sci. Acad., 67: 207-221.

Martinel, N., C. Piciarelli, C. Micheloni and G.L. Foresti, 2015. A structured committee for food recognition. Proceedings of the IEEE International Conference on Computer Vision, December 7-13, 2015, IEEE, New York, USA., pp: 92-100.

Maruyama, Y., G.C.D. Silva, T. Yamasaki and K. Aizawa, 2010. Personalization of food image analysis. Proceedings of the 2010 16th International Conference on Virtual Systems and Multimedia (VSMM), October 20-23, 2010, IEEE, New York, USA., ISBN:978-1-4244-9027-1, pp: 75-78.

Matsuda, Y., H. Hoashi and K. Yanai, 2012. Recognition of multiple-food images by detecting candidate regions. Proceedings of the 2012 IEEE International Conference on Multimedia and Expo (ICME), July 9-13, 2012, IEEE, New York, USA., ISBN:978-1-4673-1659-0, pp: 25-30.

Pouladzadeh, P., S. Shirmohammadi and R. Al-Maghrabi, 2014. Measuring calorie and nutrition from food image. IEEE. Trans. Instrum. Meas., 63: 1947-1956.

Pouladzadeh, P., S. Shirmohammadi and T. Arici, 2013. Intelligent SVM based food intake measurement system. Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), July 15-17, 2013, IEEE, New York, USA., ISBN:978-1-4673-4702-0, pp: 87-92.

Rajesh, R. and J.S. Prakash, 2011. Extreme learning machines a review and state of the art. Intl. J. Wisdom Based Comput., 1: 35-49.

Savakar, D., 2012. Identification and classification of bulk fruits images using artificial neural networks. Intl. J. Eng. Innovative Technol., 1: 35-40.

Shroff, G., A. Smailagic and D.P. Siewiorek, 2008. Wearable context-aware food recognition for calorie monitoring. Proceedings of the 12th IEEE International Symposium on Wearable Computers, September 28-October 1, 2008, IEEE, New York, USA., ISBN:978-1-4244-2637-9, pp: 119-120.

Vartak, A.P. and V. Mankar, 2013. Colour image segmentation a survey. Intl. J. Emerging Technol. Adv. Eng., 3: 681-688.

Wang, Y., Y. He, F. Zhu, C. Boushey and E. Delp, 2015. The Use of Temporal Information in Food Image Analysis. In: New Trends in Image Analysis and Processing- ICIAP 2015 Workshopspp, Murino, V., E. Puppo, D. Sona, M. Cristani and C. Sansone (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-23221-8, pp: 317-325.

Wu, W. and J. Yang, 2009. Fast food recognition from videos of eating for calorie estimation. Proceedings of the IEEE International Conference on Multimedia and Expo, June 28-July 3, 2009, IEEE, New York, USA., ISBN:978-1-4244-4290-4, pp: 1210-1213.

Yang, S., M. Chen, D. Pomerleau and R. Sukthankar, 2010. Food recognition using statistics of pairwise local features. Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 13-18, 2010, IEEE, New York, USA., ISBN:978-1-4244-6985-7, pp: 2249-2256.

Zhu, F., M. Bosch, N. Khanna, C.J. Boushey and E.J. Delp, 2015. Multiple hypotheses image segmentation and classification with application to dietary assessment. IEEE. J. Biomed. Health Inf., 19: 377-388.0345.