Food Recognition and Leftover Estimation for Daily Diet Monitoring

Gianluigi Ciocca, Paolo Napoletano(✉), and Raimondo Schettini

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione),
Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
{ciocca,napoletano,schettini}@disco.unimib.it

Abstract. Here we propose a system for automatic dietary monitoring of canteen customers based on robust computer vision techniques. The proposed system recognizes foods and estimates food leftovers. Results achieved on 1000 customers of a real canteen are promising.

Keywords: Food recognition · Leftover estimation · Diet monitoring

1 Introduction

Automatic food recognition is an important task to support the user in his daily dietary monitoring. Nowadays, technology can support the users in keep tracks of their food consumption in a more user friendly way allowing for a more comprehensive daily dietary monitoring. Recent findings showed that computer vision techniques can help to automatically recognize food [17] and estimate its quantity [25].

The works that tackle the problem of food recognition are based on different classification strategies. For example, [14] uses a k-NN classifier on local and global features; in [11] a vocabulary is constructed on textons and the images are classified using SVM; the same classifier is used in [22] where local binary pattern and relationship between SIFT interest points are used to code the local and spatial information. SVM, Artificial Neural Networks and Random Forest classification methods are evaluated in [2] on 5000 food images organized into 11 classes described in terms of different bag-of-features. Recently, deep learning algorithms are receiving great attention due to their efficiency in dealing with, and solving complex problems. Convolutional neural network (CNN) belongs to this kind of algorithms and have been successfully used in [15] and [16] on large food image datasets. Most of food recognition works, exploits only the information derived from the picture itself. A different approach is described in [4] where the context of where the picture was taken is also exploited.

Food quantity estimation is very important in the context of a dietary monitoring since on it depends the assessment of the food intakes. One of the first work that presents a system that recognizes each food item on the plate and then estimates its quantity is [25]. The problem is tackled using binary classifiers for

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 334–341, 2015.
DOI: 10.1007/978-3-319-23222-5_41

recognition, and 3D reconstruction to measure the food volumes. A calibrated camera and a calibration card are used in [26] for food recognition and portion estimation. Also the TADA dietary assessment system [21] uses a token (checkboard in this case) for food quantity estimation. Instead of using an auxiliary token, the size of the thumb is used in [23], [24], and [27]. 3D information is often exploited. For example, 3D template matching is used in [6] and [13] while 3D shape reconstruction is used in the already cited [25] and in [13]. In the latter, 3D shape reconstruction is used for food with regular shape, while area-based weight estimation is used for food without regular shape.

Food recognition and quantity estimation algorithms are used in more general, often mobile, systems for dietary monitoring purposes [2]. Examples of such systems are FoodLog [19], DietCam [20], Menu-Match [3], and FoodCam [18]. FoodLog is an online system that relies on the users to take images of their eating occasions using a camera and then send the images to a system to be processed. DietCam is a mobile application that is able to recognize foods and automatically calculate calorie content of the meal on a server. Other mobile applications are the ones presented in [28], and [1]. FoodCam is expressly designed to perform real-time food recognition on the device using efficient image representation and fast classification. Very few works consider the problem of leftover estimation. Often the problem is treated as a special case of the problem of food recognition and quantity estimation [23,28].

2 Proposed System

In this paper, we propose a system for automatic dietary monitoring of canteen customers that is based on robust computer vision techniques for automatic food recognition and leftover estimation in a canteen scenario. Although the canteen scenario includes some apparent simplifications, such as controlled image acquisition conditions, known weekly menu etc., the problem of food recognition and leftover estimation is still a challenging problem due to the enormous variations in the tray and plate composition. The visual appearance of the same dish may greatly change depending on how it is placed on the plate. Our system is able to identify and recognize the food category and estimate the amount of food in each plate. Moreover, it is designed to evaluate the leftovers as well as allowing an estimation of the truly consumed foods. Food consuption is associated to the user identity through the use of visual marker on their mobile phone. This information can be stored in more general profiling system to keep track of the user's food consumption over several days with a minimum inconvenience.

Figure 1 illustrates the user interaction and data processing involved in our system. Specifically there are seven steps:

1. The user starts its mobile application showing his personal identification number coded into a visual marker. The mobile phone is placed on the tray and the user choses the dishes placing them on the tray.

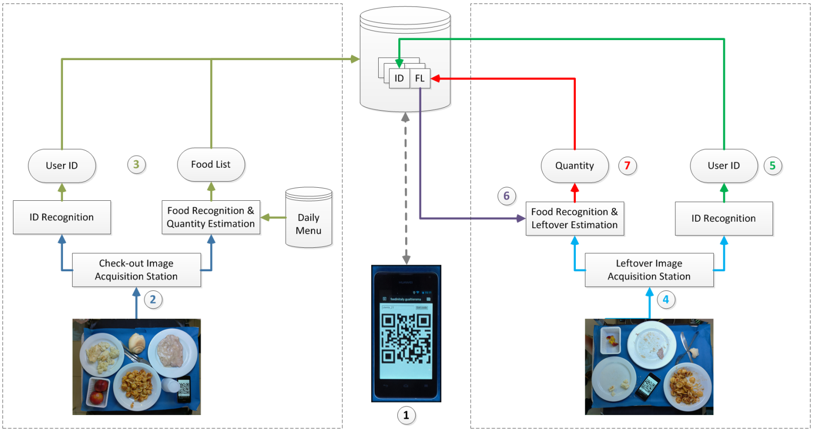
ID
FL
User ID
3
Food List
ID Recognition
Food Recognition & Quantity Estimation
Daily Menu
Check-out Image Acquisition Station
2
1
6
Quantity
7
User ID
5
Food Recognition & Leftover Estimation
ID Recognition
Leftover Image Acquisition Station
4

Fig. 1. The seven steps of our food recognition and leftover estimation system for automatic dietary monitoring of canteen customers. All the processing is carried out as a server application. The user ID is embedded into a visual marker displayed on his mobile phone.

2. Once the tray reaches the check-out acquisition station, an image of the tray is taken using a camera placed above the tray. The image is sent to the server application.
3. The image is processed. The user ID is decoded from the visual marker and the food recognition phase starts. The food recognition module detects each plate and the food placed in it is recognized according to the daily menu. The obtained food list is stored on a database into the user's dietary profile.
4. After the meal, an image of the tray containing eaten foods is taken at the leftover acquisition station and sent to the server for processing.
5. First the user ID is decoded and then used to retrieve the previous stored food list from the database.
6. The food list is used as input to the leftover estimation module that first performs food recognition. With respect to the Step 3, different processing and parameters of the food recognition are used in this step.
7. The information about the leftovers, and thus about the eaten foods, is finally logged back in the database.

As it can be seen from Figure 1, the link between the food recognition phase and the leftover estimation phase is done via the QR code representing the user ID. Its detection and recognition are performed immediately after the image acquisition through the ZBar[1] library.

2.1 Food Recognition

The overall processing pipeline is shown in Figure 2, and works as follows. To speed up the processing while maintaining enough information, the input images

[1] http://zbar.sourceforge.net/

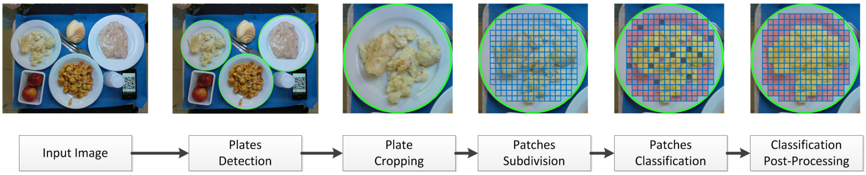

Fig. 2. Patch-based food recognition processing pipeline.

are sub-sampled to 1024 × 768 pixels. Since we are interested only on the food placed on the plates, plate detection is performed using the Hough transform for circles, with suitable parameters. To cope with the problem of having multiple foods on the same plate, we designed a patch-based food recognition algorithm. Each plate is cropped and subdivided into patches and each patch is classified into one of the candidate foods. The size of the patches influences the classification accuracy. Small patches increase the number of misclassification between patches of different classes, while large patches make the classification too noisy since non relevant information could be included in every single patch. After having experimented with different patch sizes, we found a tradeoff and we set the size to 40 × 40 pixels. From each patch, a visual descriptor is extracted and submitted to a pre-trained k-NN classifier in order to receive a classification label. The labels of the the patches are then post-processed to remove spurious labels in order to have more homogeneous groups of labels that correspond to the food regions.

We have experimented with several descriptors listed in Table 1 from different classes of approaches, such as color based, statistical, spatial-frequency or spectral, structural and hybrid [5,7,9,10]. Detailed results are reported in Section 3. It must be noted that the sample space of the classifier is constructed such that its scope is limited to samples taken from foods belonging to the daily menu list. Among the samples we have also included patches of non-food items such as plates, tablecloth, cutlery, etc. The list of recognized foods is assembled in a food list that is stored into the user's profile. Along with the food identities, the food quantities are also assessed and stored to be used for leftover estimation. These quantities, that represent reference baselines, are determined by counting the number of patches of each food.

2.2 Leftover Estimation

Within the canteen scenario, the personnel is bounded to follow the regulations provided by nutritionists in the form of nutritional tables, and to serve a specific amount of food (that depends on its calories and nutrients). This somewhat simplifies the problem of the estimation of the food quantity. In this scenario, it is more important to precisely identify what kind of foods have been chosen by the users in a given day, how much of these foods have been eaten, and to compare them with the user's dietary recommendations.

Table 1. Visual descriptors tested in our system

Name	Description	Length
CEDD	Color and Edge Directivity Descriptor	144
Gabor	Gabor features. Mean and st.dev. of RGB DFT at (θ,f)=(4,4)	96
OG	Opponent Gabor. Gabor on iter-intra channel combinations	264
LBP	Non-uniform, invariant Local Binary Pattern with (r,n)=(1,8)	54
LCC	Local Color Contrast	499
CM	Two sets of five normalized Chromaticity Moments	10
CWT	Complex Wavelet features. RGB mean and st.dev. at three scales	18

The processing pipeline starts with the user identification that is done as in the previous section. After that, the list of the daily food taken by the user is retrieved from the server database and is used to match plates before and after meal. Such a food list is also used to limit the search space of the leftover estimation. The estimation of the leftover quantity is performed by counting the food patches of each food. For a given customer (i) and a given food class (c), we define the ratio (r_{ic}^{est}). This is the ratio between the number of patches found by the leftover estimation module and the number of patches previously found by the recognition module. This is an estimation of the amount of eaten food c by the customer i:

$$r_{ic}^{est} = \frac{\#Patches\ leftover}{\#Patches\ before}.$$

Once we have identified this ratio, the corresponding amount of calories is deduced by the precompiled nutritional tables.

3 Experiments

We experimented our system in a real scenario. We monitored and recorded the meal of 1000 customers of a real canteen that corresponded to 2000 tray images (1000 before and 1000 after the meal). Each customer selected 3 dishes from the daily menu that included 15 different dishes. The images have been acquired through an automatic photographic system that includes a raspberry motherboard, an embedded camera and a motion sensor. The system automatically detects when the tray has to be acquired. All the 2000 images have been manually annotated in order to provide the ground-truth for both recognition and leftover estimation. The annotations have been created using the IAT - image annotation tool [8], that permitted to draw a polygon around the food. For a customer i, the assesment of the leftover quantity of a food c is obtained as the ratio between the areas of annotated polygons before and after the meal (r_{ic}^{gt}).

We have selected 300 customers (600 tray images) to train the system and 700 for testing. The training set has been selected such that all the 15 dishes were equally represented. The training step has been performed with a k-NN algorithm with $k = 7$. The food distribution of the 700 test customers is reported in Table 2. As can be seen, the food classes are not uniformly distributed

To cope with the class imbalance problem of the test set we jointly used two assessment metrics for food recognition: the *Standard Accuracy (SA)* and the

Table 2. Food recognition rate of all the visual descriptors considered. Best performance are reported in bold.

			visual descriptors						
Classes	w_c(%)		CEDD	OG	Gabor	LBP	LLC	CM	CWT
bistecca	(3.8%)		100.00	100.00	100.00	27.50	97.50	91.25	80.00
carote	(7.6%)		100.00	100.00	100.00	100.00	100.00	98.75	100.00
cavolfiore	(8.6%)		100.00	100.00	98.89	97.22	98.33	97.22	98.33
fagiolini	(7.6%)		100.00	100.00	100.00	99.38	100.00	100.00	96.25
frittata	(7.6%)		100.00	100.00	100.00	81.25	93.75	83.12	100.00
fusilli ragu	(8.6%)		100.00	100.00	100.00	85.56	100.00	97.22	100.00
insalata mista	(2.4%)		100.00	92.00	42.00	58.00	100.00	90.00	32.00
lenticchie	(7.1%)		98.67	99.33	96.67	68.00	94.67	28.67	57.33
minestra	(6.7%)		100.00	100.00	97.86	99.29	97.86	93.57	100.00
pasta cime rapa	(8.6%)		100.00	100.00	100.00	100.00	100.00	100.00	100.00
pasta sugo	(2.4%)		100.00	100.00	100.00	28.00	76.00	100.00	98.00
piselli	(7.1%)		99.33	100.00	98.67	94.67	100.00	88.00	98.00
pollo ferri	(7.6%)		96.86	97.48	67.30	62.26	76.10	93.71	69.18
scaloppina	(8.6%)		98.90	99.45	99.45	13.81	98.34	97.79	98.90
tortino	(5.7%)		91.67	90.83	79.17	22.50	79.17	83.33	80.00
		SA	**99.05**	99.00	94.33	74.14	95.05	89.57	90.38
		MAA	**99.03**	98.61	92.00	69.16	94.11	89.51	87.20

Macro Average Accuracy (MAA) [12]. Denoting NP_c the number of positives, i.e., the number of times the class c occurs in the dataset; TP_c the number of *true positives* for class c, i.e., the number of times that the system recognizes the dish c; C the number of classes, for each class, the metrics can be defined as follows:

$$SA = \frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C} NP_c}; \qquad MAA = \frac{1}{C}\sum_{c=1}^{C} A_c = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{NP_c}.$$

Regarding the food recognition task, the system showed a very high performance based on both assessment metrics, see Table 2. The *CEDD* descriptor achieved the highest accuracy while the LBP and CWT achieved the lowest accuracy. Concerning the evaluation of the leftover estimation module, we measured the overall relative error ($Error$) as:

$$Error = \sum_{c=1}^{C} w_c \sum_{i=1}^{I} |r_{ic}^{gt} - r_{ic}^{est}|,$$

where w_c is the class weight and I is the number of test customers. The class weight is defined as the number of elements of the class divided by the total number of elements. The system is capable of estimating the relative quantity of eaten food with an average error of about 15 percentage points, with the best and worst cases being 7 and 34 percentage points respectively.

4 Conclusions

The proposed food recognition and leftover estimation system can serve multiple purposes: first, at the check-out station, the food recognition allows to keep track the eaten food and the user's dietary habits; second, using the list of recognized foods, an automatic billing procedure can be activated speeding up the check-out; third, by evaluation the leftovers, we can better estimate the food intakes in terms of calories ingested. Results achieved on a real canteen scenario are promising with an average accuracy in recognition of about 99%, and and average error in food estimation of 15 percentage points.

Acknowledgments. This work is part of the research Project titled Feedin' Italy. The project is co-funded by European Union, The italian government, and Regione Lombardia under CE regulations 1083/2006 and 1828/2006.

References

1. Ahmad, Z., Khanna, N., Kerr, D.A., Boushey, C.J., Delp, E.J.: A mobile phone user interface for image-based dietary assessment. In: IS&T/SPIE Electronic Imaging, p. 903007. International Society for Optics and Photonics (2014)
2. Anthimopoulos, M.M., Gianola, L., Scarnato, L., Diem, P., Mougiakakou, S.G.: A food recognition system for diabetic patients based on an optimized bag-of-features model. IEEE Journal of Biomedical and Health Informatics **18**(4), 1261–1271 (2014)
3. Beijbom, O., Joshi, N., Morris, D., Saponas, S., Khullar, S.: Menu-match: restaurant-specific food logging from images. In: 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 844–851. IEEE (2015)
4. Bettadapura, V., Thomaz, E., Parnami, A., Abowd, G., Essa, I.: Leveraging context to support automated food recognition in restaurants. In: 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 580–587 (2015)
5. Bianconi, F., Harvey, R., Southam, P., Fernández, A.: Theoretical and experimental comparison of different approaches for color texture classification. Journal of Electronic Imaging **20**(4) (2011)
6. Chae, J., Woo, I., Kim, S., Maciejewski, R., Zhu, F., Delp, E.J., Boushey, C.J., Ebert, D.S.: Volume estimation using food specific shape templates in mobile image-based dietary assessment. In: IS&T/SPIE Electronic Imaging, p. 78730. International Society for Optics and Photonics (2011)
7. Chatzichristofis, S.A., Boutalis, Y.S.: CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 312–322. Springer, Heidelberg (2008)
8. Ciocca, G., Napoletano, P., Schettini, R.: Iat-image annotation tool: Manual. arXiv preprint arXiv:1502.05212 (2015)
9. Cusano, C., Napoletano, P., Schettini, R.: Intensity and color descriptors for texture classification. In: IS&T/SPIE Electronic Imaging, p. 866113. International Society for Optics and Photonics (2013)
10. Cusano, C., Napoletano, P., Schettini, R.: Combining local binary patterns and local color contrast for texture classification under varying illumination. JOSA A **31**(7), 1453–1461 (2014)
11. Farinella, G., Moltisanti, M., Battiato, S.: Classifying food images represented as bag of textons. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5212–5216 (2014)
12. He, H., Ma, Y.: Imbalanced Learning: Foundations, Algorithms, and Applications. John Wiley & Sons (2013)

13. He, Y., Xu, C., Khanna, N., Boushey, C., Delp, E.: Food image analysis: segmentation, identification and weight estimation. In: 2013 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2013)
14. He, Y., Xu, C., Khanna, N., Boushey, C., Delp, E.: Analysis of food images: features and classification. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 2744–2748 (2014)
15. Kagaya, H., Aizawa, K., Ogawa, M.: Food detection and recognition using convolutional neural network. In: Proceedings of the ACM International Conference on Multimedia, MM 2014, pp. 1085–1088 (2014)
16. Kawano, Y., Yanai, K.: Food image recognition with deep convolutional features. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2014 Adjunct, pp. 589–593 (2014)
17. Kawano, Y., Yanai, K.: Foodcam-256: a large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights. In: Proceedings of the ACM International Conference on Multimedia, MM 2014, pp. 761–762 (2014)
18. Kawano, Y., Yanai, K.: Foodcam: A real-time food recognition system on a smartphone. Multimedia Tools and Applications, 1–25 (2014)
19. Kitamura, K., Yamasaki, T., Aizawa, K.: Foodlog: capture, analysis and retrieval of personal food images via web. In: Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities, pp. 23–30 (2009)
20. Kong, F., Tan, J.: Dietcam: Automatic dietary assessment with mobile camera phones. Pervasive and Mobile Computing **8**(1), 147–163 (2012)
21. Mariappan, A., Bosch, M., Zhu, F., Boushey, C.J., Kerr, D.A., Ebert, D.S., Delp, E.J.: Personal dietary assessment using mobile devices, vol. 7246, pp. 72460Z-1–72460Z-12 (2009)
22. Nguyen, D.T., Zong, Z., Ogunbona, P.O., Probst, Y., Li, W.: Food image classification using local appearance and global structural information. Neurocomputing **140**, 242–251 (2014)
23. Pouladzadeh, P., Shirmohammadi, S., Al-Maghrabi, R.: Measuring calorie and nutrition from food image. IEEE Transactions on Instrumentation and Measurement **63**(8), 1947–1956 (2014)
24. Pouladzadeh, P., Villalobos, G., Almaghrabi, R., Shirmohammadi, S.: A novel svm based food recognition method for calorie measurement applications. In: 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 495–498 (2012)
25. Puri, M., Zhu, Z., Yu, Q., Divakaran, A., Sawhney, H.: Recognition and volume estimation of food intake using a mobile device. In: 2009 Workshop on Applications of Computer Vision (WACV), pp. 1–8 (2009)
26. Sun, M., Liu, Q., Schmidt, K., Yang, J., Yao, N., Fernstrom, J., Fernstrom, M., DeLany, J.P., Sclabassi, R.: Determination of food portion size by image processing. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, pp. 871–874 (2008)
27. Villalobos, G., Almaghrabi, R., Pouladzadeh, P., Shirmohammadi, S.: An image processing approach for calorie intake measurement. In: 2012 IEEE International Symposium on Medical Measurements and Applications Proceedings, pp. 1–5 (2012)
28. Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C., Ebert, D., Delp, E.: The use of mobile devices in aiding dietary assessment and evaluation. IEEE Journal of Selected Topics in Signal Processing **4**(4), 756–766 (2010)