



# HHS Public Access

Author manuscript

*J Health Med Inform.* Author manuscript; available in PMC 2018 August 10.

Published in final edited form as:

*J Health Med Inform.* 2017 ; 8(3): . doi:10.4172/2157-7420.1000272.

## A Survey on Automated Food Monitoring and Dietary Management Systems

Vieira Bruno, Silva Resende, and Cui Juan\*

Department of Computer Science and Engineering, University of Nebraska–Lincoln, NE, USA

### Abstract

Healthy diet with balanced nutrition is key to the prevention of life-threatening diseases such as obesity, cardiovascular disease, and cancer. Recent advances in smartphone and wearable sensor technologies have led to a proliferation of food monitoring applications based on automated food image processing and eating episode detection, with the goal to conquer drawbacks of the traditional manual food journaling that is time consuming, inaccurate, underreporting, and low adherent. In order to provide users feedback with nutritional information accompanied by insightful dietary advice, various techniques in light of the key computational learning principles have been explored. This survey presents a variety of methodologies and resources on this topic, along with unsolved problems, and closes with a perspective and boarder implications of this field.

### Keywords

Food monitoring; Food image classification; Food image dataset; Automatic nutrient assessment; Machine learning

### Introduction

Many people face challenges to maintain healthy diet and manage their weight these days, while knowing bad eating habits lead to overweight and obesity that increase the risk of heart diseases, hypertension, other metabolic comorbidities such as type 2 diabetes, and cancer [1]. Personal diet management is always warranted in these scenarios, which often involves manual food logging that is time consuming and tedious [2]. By virtues of growth of smartphone use, several mobile applications have been developed to facilitate food journaling, such as MyFitnessPal, LoseIt and Fooducate, and many have demonstrated great potential in effective diet control [3]. For example, a study shows higher user retention with smartphone-based diet logging compared to the websites and paper diary in a period of six months [4]. Teenagers are willing to take food images using a mobile food recorder before eating [5]; and the dietary feedback contributes to weight loss [6]. However, many of these applications require significant manual input from users and suffer from the low

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

\*Corresponding author: Juan Cui, Systems Biology and Biomedical Informatics (SBBI) Laboratory, Department of Computer Science and Engineering, University of Nebraska–Lincoln, NE, 68588, USA, Tel: 4024725023, jcui@unl.edu.

performance in assessing the exact ingredients and food portion [7], which has hindered the long-term use from user.

The current consensus objective on this topic is to develop new methods that can automatically identify food items and estimate nutrients based on food images, utilizing cutting-edge techniques in Computer Vision and Machine Learning and ideally being friendly, effort free and accurate for user to keep track of their meals. Along this line of research, several key issues have been raised including the following. First, the food image databases are expected to be comprehensive, containing large number of food classes to cover the food diversity and abundant images per class to reflect the food image discrepancy when training a classification system [8]. Second, reliable food segmentation is highly recommended to identify all possible items in an image and separate them from the background regardless the lighting conditions or if the food are mixed or not [9]. Subsequently, classification will be performed on each segmented item using machine learning models that are trained based on large food datasets. Last, volume and weight estimation can be performed on each identified item, followed by the nutrient assessment [10-12]. The workflow of an automated food monitoring system that connects these components is presented in Figure 1. It is notable that every aforementioned step involves technical challenges, e.g., it is difficult to estimate food volume based on 2-dimensional images.

In addition to the image-based strategy, several wearable devices, such as glasses with load cells [13] or connected to sensors on temporalis muscle and accelerometer [14] and wrist motion track [15], have been explored to detect food intake events automatically. The collected information about eating episodes, pertinent to users' diet habit pattern, can serve as starting point for food consumption analysis and diet interventions, e.g., providing user recommendations for physical exercise, healthier food, or eating habit [16,17].

In this paper, we review the most relevant applications on automatic food monitoring (till April 2017) that focus on addressing each aforementioned challenge. We specifically introduce current food image databases in section 2, followed by a survey on next section existing methods for segmentation, feature extraction, classification, and volume and nutrient estimation. In addition, a few studies on food-monitoring wearable devices and diet invention are depicted, respectively. Finally, we close the review by discussing the remaining challenges and presenting future outlook in this field.

## Food Image Databases

A comprehensive collection of quality food images is key to train a food-classification model and benchmark the prediction performance, i.e., a common procedure to verify if a new classifier outperforms previous methods is to compare their classification performance on large food image databases such as Food-101 [18], UEC Food-100 [9], and UEC Food-256 [19]. Current food image datasets vary in many aspects, such as, type of cuisine, number of food groups, and total images per food class. For instance, Menu-Match dataset [20] contains 41 food classes and a total of 646 images captured in 3 distinct restaurants

while PFID [21] has 61 classes with a total of 1098 pictures captured in fast food restaurants and laboratory. Table 1 gives a summary of different databases with their respective features.

It is noticeable that there is no benchmark food image database for general classification purpose since most databases archive specific food type. For examples, the UNIMIB2016 database [8] has Italian food images from a campus dining hall and the UEC Food-100 [9] consists of items from Chinese culinary. Similarly, Chen [22] and PFID [21] consist of images from traditional Japanese dishes and American fast food, respectively. On the other hand, Food-101 [18] and UEC Food-256 [19] contain a mix of eastern and western food. Except for food type, other image features such as if the picture was obtained in the wild, in a controlled environment, or whether the image is segmented or not has been taken into consideration when developing those databases (Table 1).

## Image Based Food Recognition

### Food image segmentation

Segmentation is an important process to separate parts of a scene. When dealing with food, the objective is to localize and extract food items from the image [23-26]. It takes place before food classification when authors attempt to identify multiple food items in the image [8,27] or estimate volume [11,12], which often contributes to improved classification accuracy [9,12].

It is challenging to segment food images since they may not present specific attributes such as edges and defined contour [28]. Food items can be on top of each other or being obstructed by another component, making it hidden in the given image [28]. Meanwhile, external factors such as illumination can interfere negatively in this step, where shadows can be identified as part of the food or even a new food item [12,29].

Several methods have been proposed to address the segmentation issue, summarized in Table 2. For examples, one asks user to draw bounding boxes over food items on the smartphone screen, and performs segmentation using GrabCut algorithm over selected areas [27]. Another segments items by integrating four methods to detect candidate region, including the whole image (assuming each image has one food), Deformable Part Model (DPM, a method utilizing sliding windows to detect object regions), circle detector (detecting circular in an image), and JSEG segmentation to segment regions [9]. A similar approach in Ciocca et al. [8] combined different strategies including image saturation, binarization, JSEG segmentation, and morphological operations (noise removal) to segment multiple food items. In addition, the work presented in Yang et al. [28] tries to segment food by its ingredients and their spatial relationship applying Semantic Texton Forest (STF).

Of particular interest is that Deep Learning approach has been used to tackle food segmentation [11,30], although at its early stage. For example, the application named Im2Calories utilized the Convolution Neural Network (CNN) model that provides unary potentials of a conditional random fields and a fully connected graph to perform edge-sensitive label smoothing [11], which increased the overall classification accuracy (Table 2).

## Feature extraction

Image objects can be recognized based on their characteristics, such as colour, shape and texture [31]. According to Hassannejad et al. [32], selection of relevant features is important when building a recognition model capable of identifying food items. General image features, as mentioned above, may not be descriptive enough to distinguish foods since the properties of the same good may change when the food is prepared in different ways [23]. For example, Penne and Spaghetti have same colour and texture but distinct shape.

In order to extract informative visual information from food image, descriptors such as Local Binary Patterns (LBP), color information, Gabor filter, Scale-Invariant Feature Transform (SIFT) [22], called handcrafted features, can be applied (illustrated in Table 3). Different features and their fusion often result in different classification performance. For instance, when SIFT and LBP were used individually on Chen dataset [22], it achieves accuracy of 53% and 45.9%, respectively; when they were combined with additional colour and Gabor filter, accuracy rises to 68,3%. Based on the same dataset, another study, Menu-Match [20], extracted the SIFT, LBP and colour in different settings, along with HOG and MR8 and obtained the accuracy of 77.4%. It also illustrates how sensitive a classification can be when the same feature is extracted but with different parameters.

## Food classification

Currently, there are two major classification strategies for food image recognition: 1) Traditional machine learning-based approach using handcrafted features and 2) Deep Learning-based approach. The former usually start with a set of visual features extracted from the food image and use them to train a prediction model based on Machine Learning algorithms such as Support Vector Machine [20], Bag of Features [31], or K Nearest Neighbors [8]. In contrast, emerging deep learning architectures have a large number of connected layers that are able to learn features, followed by a final layer responsible for classification [33]. Recent approaches based on Deep Learning become more popular and effective, e.g., the study in Christodoulidis et al. [34] obtained astonishing results in the ImageNet's Large Scale Visual Recognition Challenge 2012 (ILSVRC2012).

The following example compares a classifier trained with handcrafted features with a deep learning architecture. In Yanai and Kawano [35], color and HOG features are classified using a similar strategy to Bag of Features, called Fisher Vectors, which achieved accuracy of 65.3% on UEC Food-100 [9]. On the same database, the Deep Learning architecture DCNN-FOOD [35] was created and showed an improvement of 13.5% over the handcrafted method. A major advantage of Deep Learning method is that they can learn relevant features from images automatically, which is particularly important in the cases when the pre-defined features are not discriminative enough [32]. More studies based on both methods are shown in Table 3. Clearly, a common issue with most current methods is that the performance was presented mainly based on overall accuracy where the assessment of sensitivity and specificity was missing (Table 3).

## Food volume estimation and nutrient analysis

After identifying all food items from an image, it is important to assess the nutrients included, e.g., the carbohydrates, sugar, or total calorie, which will require volume/weight estimation, another challenge. In fact, not even an expert dietitian can estimate the total calories without a precise instrument, e.g., a scale. Taking image-based calorie estimation as an example: first, food candidate regions must be recognized, segmented, and classified correctly [22,36]; the volume from each segmented item will be calculated; and the nutrient can be estimated based on a nutritional facts table [37-39], such as USDA Food Composition Database [40].

The most challenging part is to estimate food's volume from 2-dimensional image which normally does not have the depth information, unless reference objects are placed next to the meal [8,41]. Volume can be underestimated or overestimated with interference from external factors, such as lighting conditions, blurred images, and noisy background [22], only few strategies were reported for estimation of food volume and calorie intake as currently the major focus in this domain still lies in the food classification (Table 4) [32].

As listed in Table 4, crowdsourcing [42] and a depth sensor camera [11,22] have been utilized for food volume estimation and nutrition assessment. Although leading to promising results, these studies were conducted either in a controlled environment or using an extra camera that is not practical in real-world events. In addition, user's finger was also used as reference when one takes a picture from the top and side views of the plate to estimate food volume [12]. The concern here is that multiple food items overlap in the side view, making it hard to distinguish. Similarly, another reference object, a checkerboard, was used to help obtaining depth information alongside camera calibration [10], which also needs users to carry additional equipment in order to estimate food's volume.

Note that those methods are mostly tied to a controlled environment. For example, it has stated that a broader study outside the laboratory is not feasible because nutrient values vary depending on how the food was prepared and there is no broad nutritional database for prepared foods yet [11]. On the other hand, it performed volume estimation for only 7 items in Woo et al. [43], while the study only matched classified food to annotated menu items with respective (Table 4) known calories in Beijbom et al. [20].

## Wearable Device-Based Food Monitoring

Other than monitoring food intake through image processing, several wearable devices have been developed for auto-detection of eating episodes. For example, a proof of concept called Glassense [13] utilizes a pair of glasses with load cells to detect user's digestive behaviours through facial signals. Likewise, glasses connected to a sensor placed on the temporalis muscle and an accelerometer was also presented to detect food intake when users are physically active and/ or talking [14]. In addition, a wrist motion tracker was developed to identify eating activities and measure food intake [15].

Although these approaches can detect eating activities with decent resolution, more follow-up research efforts are needed to explore the relationships between eating activities and nutrient intake and calories consumption.

## Diet Intervention

Dietary intervention can be realized after the aforementioned diet management systems learn adequate information about the individual's eating habits. Often it requires functionality similar to a diet advisor capable of giving users feedbacks to improve their health, e.g., eat less often or replace A by B in the meal for weight loss [6]. Recent applications are more sophisticated in this regard. For examples, Faiz et al. [16] introduces a Semantic Healthcare Assistant for Diet and Exercise (SHADE) that can identify user habits and generate suggestions not only for diet, but also for exercise for diabetic control. Similarly, Lee et al. [17] presents a personal food recommendation agent that can create a meal plan according to a person's lifestyle and particular health needs towards a certain health goal.

## Remaining Challenges

As mentioned above, despite of the advances in food recognition technologies, there are remaining challenges with respect to each analytical step. For example, food image datasets and classification methods are highly related since the former provide training data for the latter. Current image databases tend to grow in number of classes to incorporate different types of food, as what happened to Food201 Segmented [11], Food85 [25], and UEC Food-256 [19]. Meanwhile, classifiers are developed based on new architecture that is capable of identifying new food items. Since the Deep Learning approaches can provide better classification accuracy when trained on larger datasets [33], there is a possible also a need to generate more food images from existing datasets by randomly cropping images and apply distortions like brightness, contrast, saturation and hue [32].

Although segmentation of food items has shown significant improvements in Zhu et al. [10], it is still difficult to segment hidden food item and mixed food. Other factors such as lightning can also contribute negatively to segment foods. For example, shadows can be considered as part of food or candidate regions by algorithms. Methods based on manually-selected candidate items can be promising [30], however, the bounding box size may be influential [27].

Nutrient and calorie estimation remains the most challenging problem in automated diet monitoring systems since it is highly dependent on food segmentation and volume estimation [11]. Undoubtedly, calories can be overestimated or underestimated if any of the other steps is erroneous. However, as discussed above, volume estimation based on 2D images are still far from satisfactory even using the effective reference objects such as a checkerboard [43] and finger [44]. Note this problem can be solved by using stereo cameras, as illustrated in im2Calories [11], which requires extra accessories, or using SmartPlate, a device that integrates multiple scales into a dining plate to weight food items. Obviously, once all those new functionalities and sensors are embedded in the smartphones, all such complexity [45] of the problem can be alleviated significantly.

## Conclusion

In this review, we have surveyed a wide range of strategies in computer vision and artificial intelligence specifically designed for automated food recognition and dietary intervention. Particularly, the entire framework can be broken down into four parts that involve developments of comprehensive food image databases, classifiers capable for food item recognition, and strategies for food volume estimation, nutrient analysis that provide information for diet intervention. Even though improved performance has been demonstrated, challenging issues still remain and desire novel algorithms and techniques. Worth mentioning is the increased appreciation of using Deep Learning models for food image classification, which has outperformed traditional methodologies using handcrafted features. Increased application of wearable sensor devices, especially those can be integrated into smartphone, will revolutionize this line of research and as a whole the food monitoring system will help generate novel insights in effective health promotion and disease prevention.

## Acknowledgments

This work was supported by the National Institutes of Health funded COBRE grant [1P20GM104320] from National Institutes of Health, UNL Food For Health seed grant and the Tobacco Settlement Fund as part of Cui's startup grant.

## References

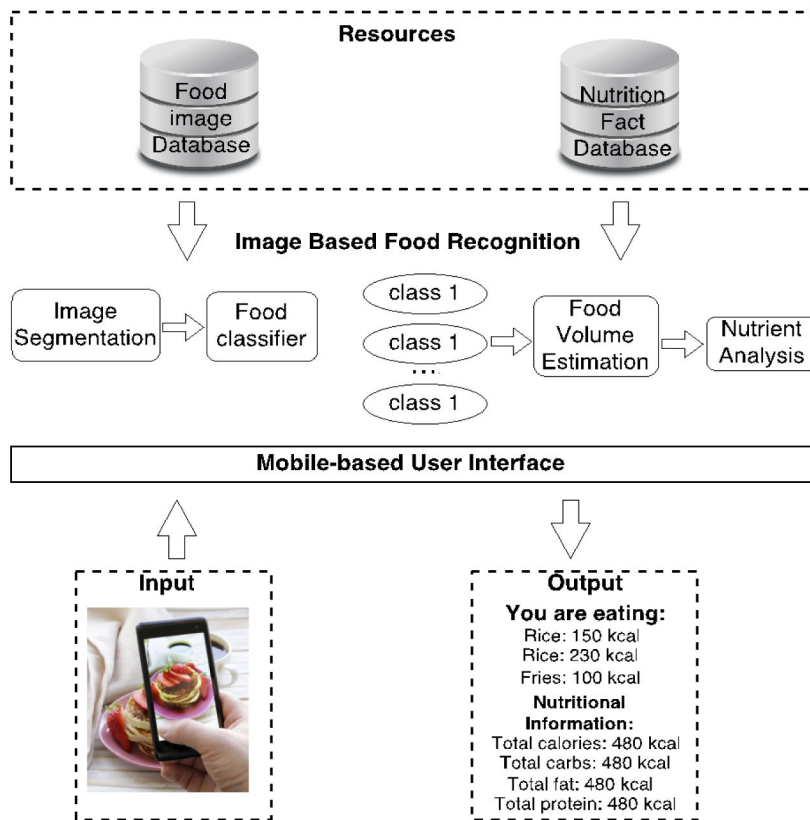
1. Must A , Spadano J , Coakley EH , Field AE , Colditz G , et al. (1999) The disease burden associated With overweight and obesity. *JAMA* 282: 1523–1529.10546691
2. Champagne CM , Bray GA , Kurtz AA , Monteiro JB , Tucker E , et al. (2002) Energy intake and energy expenditure: A controlled study comparing dietitians and non-dietitians. *J Am Diet Assoc* 102: 1428–1432.12396160
3. Haapala I , Barengo NC , Biggs S , Surakka L , Manninen P (2009) Weight loss by mobile phone: a 1-year effectiveness study. *Public Healt Nutr* 12: 2382–2391.
4. Carter MC , Burley VJ , Nykjaer C , Cade JE (2013) Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. *J Med Internet Res* 15: 32.
5. Boushey CJ , Harray AJ , Kerr DA , Schap TE , Paterson S , et al. (2015) How willing are adolescents to record their dietary intake? The mobile food record. *JMIR MHealth UHealth* 3: 47.
6. Kerr DA , Harray AJ , Pollard CM , Dhaliwal SS , Delp EJ , et al. (2016) The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support in young adults. *Int J Behav Nutr Phys Act* 13: 52.27098449
7. Cordeiro F , Epstein DA , Thomaz E , Bales E , Jagannatha AK , et al. (2015) Barriers and negative nudges: exploring challenges in food journaling. *Proc Sigchi Conf Hum Factor Comput Syst* 2015: 1159–1162.26894233
8. Ciocca G , Napoletano P , Schettini R (2016) Food recognition: a new dataset, experiments and results. *IEEE J Biomed Health Inform* 99: 1-1.
9. Matsuda Y , Hoashi H , Yanai K (2012) Recognition of multiple-food images by detecting candidate regions. *IEEE Int Confe Multimed Expo* 25: 25–30.
10. Zhu F , Bosch M , Khanna N , Boushey CJ , Delp EJ (2015) Multiple hypotheses image segmentation and classification with application to dietary assessment. *IEEE J Biomed Health Inform* 19: 377–388.25561457



11. meyers a (2015) Im2calories: Towards an automated mobile vision food diary. *IEEE Int Confe Compu Visi2015*: 1233–1241.
12. Pouladzadeh P , Shirmohammadi S , Yassine A (2014) Using graph cut segmentation for food calorie measurement. *IEEE Int Symp Med Measur Appl* 25: 1–6.
13. Chung J , Chung J , Oh W , Yoo Y , Lee WG , et al. (2017) A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Sci Rep* 7: 41690.28134303
14. Farooq M , Sazonov E (2016) A novel wearable device for food intake and physical activity recognition. *Sensors* 16: 7.
15. Dong Y , Hoover A , Scisco J , Muth E (2012) A new method for measuring meal intake in humans via automated wrist motion tracking. *Appl Psychophysiol Biofeedback* 37: 205–215.22488204
16. Faiz I , Mukhtar H , Khan S (2013) An integrated approach of diet and exercise recommendations for diabetes patients. *Qual Rep* 18: 1–22.
17. Lee CS , Wang MH , Li HC , Chen WH (2008) Intelligent ontological agent for diabetic food recommendation. *IEEE Int Conf Fuz Sys* 25: 1803–1810.
18. Bossard L , Guillaumin M , Gool LV (2014) Food-101-Mining discriminative components with random forests in computer vision. *ECCV* 2014: 446–461.
19. Kawano Y , Yanai K (2014) Automatic expansion of a food image dataset leveraging existing categories with domain adaptation in computer vision. *ECCV* 2014: 3–17.
20. Beijbom O , Joshi N , Morris D , Saponas S , Khullar S (2015) menu-match: restaurant-specific food logging from images in 2015. *IEEE* 25: 844–851.
21. Chen M , Dhingra K , Wu W , Yang L , Sukthankar R , et al. (2009) PFID: Pittsburgh fast-food image dataset in 2009. *IEEE* 1: 289–292.
22. Chen MY (2012) Automatic chinese food identification and quantity estimation *NY* 29: 1–29.
23. Chen J , Ngo C (2016) Deep-based ingredient recognition for cooking recipe retrieval. *ACM* 1: 32–41.
24. Farinella GM , Moltisanti M , Battiato S (2014) Classifying food images represented as Bag of Textons. *IEEE* 1: 5212–5216.
25. Hoashi H , Joutou T , Yanai K (2010) Image recognition of 85 food categories by feature fusion. *IEEE* 1: 296–301.
26. Mariappan A (2009) Personal dietary assessment using mobile devices. *Proc Spie* 7246: 1–22.
27. Kawano Y , Yanai K (2013) Real-Time Mobile Food Recognition System. *IEEE* 25: 1–7.
28. Yang S , Chen M , Pomerleau D , Sukthankar R (2010) Food recognition using statistics of pairwise local features. *IEEE* 1: 2249–2256.
29. He Y , Xu C , Khanna N , Boushey CJ , Delp EJ (2014) Analysis of food images: Features and classification. *IEEE* 1: 2744–2748.
30. Shimoda W , Yanai K (2015) CNN-Based food image segmentation without pixel-wise annotation. *ICIAP* 2015: 449–457.
31. Anthimopoulos MM , Gianola L , Scarnato L , Diem P , Mougiakakou SG (2014) A food recognition system for diabetic patients based on an optimized bag-of-features model. *J Biomed Health Inform* 18: 1261–1271.
32. Hassannejad H , Matrella G , Ciampolini P , Munari ID , Mordonini M , et al. (2016) Food image recognition using very deep convolutional networks. *Intl Work Multime Ass Die Manag* 1: 41–49.
33. Krizhevsky A , Sutskever I , Hinton GE (2012) Imagenet classification with deep convolutional neural networks.” *CJC* 25: 1097–1105.
34. Christodoulidis S , Anthimopoulos M , Mougiakakou S (2015) food recognition for dietary assessment using deep convolutional neural networks. *ICIAP* 2015: 458–465.
35. Yanai K , Kawano Y (2015) Food image recognition using deep convolutional network with pre-training and fine-tuning. *IEEE* 1: 1–6.
36. Liu C , Cao Y , Luo Y , Chen G , Vokkarane V , et al. (2016) Deep food: deep learning-based food image recognition for computer-aided dietary assessment. *Dig Healt* 3: 37–48.
37. <https://ndb.nal.usda.gov/ndb/>



38. Kawano Y , Yanai K (2014) FoodCam-256: A large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. *IEEE* 3: 761–762.
39. Kawano Y , Yanai K (2014) Food image recognition with deep convolutional features. *ACM* 2014: 589–593.
40. Pouladzadeh P , Kuhad P , Peddi SVB , Yassine A , Shirmohammadi S (2016) Food calorie measurement using deep learning neural network. *IEEE* 2016: 1–6.
41. Zhu F (2010) The use of mobile devices in aiding dietary assessment and evaluation. *IEEE* 4: 756–766.
42. Noronha J , Hysen E , Zhang H , Gajos KZ (2011) platemate: crowdsourcing nutritional analysis from food photographs. *ACM* 2011: 1–12.
43. Woo I , Otsmo K , Kim S , Ebert DS , Delp EJ , et al. (2010) Automatic portion estimation and visual refinement in mobile dietary assessment. *Proc SPIE Int Soc Opt Eng* 7533: 753300–753310
44. Villalobos G , Almaghrabi R , Pouladzadeh P , Shirmohammadi S (2012) An image processing approach for calorie intake measurement. *IEEE* 2012: 1–5.
45. Kumar A , Tanwar P , Nigam S (2016) Survey and evaluation of food recommendation systems and techniques. *IEEE* 2016: 3592–3596.



**Figure 1:** The workflow of an automated food monitoring system that connects various components discussed in the main text.

**Table 1:**

Food image databases.

Study	Database	Image content	Total # of class/image	Acquisition	Reference
Chen et al., 2009	PFID	Fast food items from USA	61/1098	Images taken in restaurants and in lab, with white background	[21]
Mariappan, 2009	TADA *	Common food in USA	256 food+50 replicas	Images collected in a controlled environment	[26]
Hoashi et al., 2010	Food85 *	Japanese food	85/8500	Images derived from previous database with 50 Japanese food category and web	[25]
Chen, 2012	Chen	Chinese food	50/5000	Images downloaded from the web	[22]
Matsuda et al., 2012	UEC Food-100	Popular Japanese food	100/9060	Images acquired by digital camera (each photo has a bounding box indicating the location of the food item)	[9]
Farinella et al., 2014	Diabetes	Selected food	11/4868	Images downloaded from the web	[24]
Bossard et al., 2014	Food-101	Popular food in USA	101/101000	Images downloaded from the web	[18]
Kawano and Yanai, 2014	UEC Food-256	Popular foods in Japan and other countries	256/31397	Images acquired by digital camera (each photo has a bounding box indicating the location of the food item)	[19]
Meyers, 2015	Food201-Segmented *	Popular food in USA	201/12625	Images derived from Food 101 dataset; segmented	[11]
Beijbom et al., 2015	Menu-Match	Food from three restaurants (Asian, Italian, and soup)	41/646	Images taken by authors	[20]
Ciocca et al., 2016	UNIMIB2016	Food from dining hall	73/1027	Images acquired by digital camera in dining hall; segmented	[8]
Chen and Ngo, 2016	Vireo	Chinese dishes	172/110241	Images downloaded from the web	[23]

\* Proprietary database

**Table 2:**

Food segmentation methods.

Study	Approach	Performance	Reference
Yang et al., 2010	Semantic Texton Forest calculates the probability for each pixel to belong to one of the food classes.	Output from Semantic Texton Forest is far from a precise parsing of an image	[28]
Matsuda et al., 2012	Combined techniques: whole image, DPM, circle detector and JSEG segmentation	Overall accuracy to 21% (top 1) and 45% (top 5) <sup>*</sup>	[9]
Kawano and Yanai, 2013	Each food item within user generated bound boxes is segmented by GrabCut algorithm	Performance depending on the size of the bounding boxes	[27]
Pouladzadeh et al., 2014	Graph cut segmentation algorithm to extract food items and user's finger	Overall accuracy of 95%	[12]
Shimoda and Yanai, 2015	CNN model searching for food item based on fragmented reference	Detects correct bounding boxes around food items with mean average precision of 49.9% when compared to ground truth values	[30]
Meyers, 2015	DeepLab model	Classification accuracy increases with conditional random fields	[11]
Zhu et al., 2015	Multiple segmentations generated for an image and selected by a classifier	It outperforms normalized cut	[10]
Ciocca et al., 2016	Combines saturazation, binarization, JSEG segmentation and morphological operations	Achieves better segmentation than using JSEG-only approach	[8]

<sup>\*</sup> Top 1 and/or Top 5 indicate that the performance of the classification model was evaluated based on the first assigned class with the highest probability and/or the top 5 classes among the prediction for each given food item, respectively

**Table 3:**

Traditional and deep learning classification methods.

Traditional Methods						
Study	Approach		Database	Performance **		Reference
	Features	Classifier		Top1 Acc.	Top5 Acc.	
Chen, 2012	SIFT, LBP, color and gabor	Multi-class Adaboost	Chen	68.3%	90.9%	[22]
Beijbom et al., 2015	SIFT, LBP, color, HOG and MR8	SVM		77.4%	96.2%	[20]
Anthimopoulos et al., 2014	SIFT and color	Bag of Words and SVM	Diabetes	78.0%	-	[31]
Bossard et al., 2014	SURF and L*a*b color values	RFDC	Food-101	50.8%	-	[18]
Hoashi et al., 2010	Bag of features, color, gabor texture and HOG	MKL	Food85	62.5%	-	[25]
Beijbom et al., 2015	SIFT, LBP, Color, HOG and MR8	SVM	Menu-Match	51.2% *		[20]
Christodoulidis et al., 2015	Color and LBP	SVM	Local dataset	82.2%	-	[34]
Pouladzadeh et al., 2014	Color, texture, size and shape	SVM		92.2%	-	[12]
Pouladzadeh et al., 2014	Graph Cut, color, texture, size and shape	SVM		95.0%	-	[12]
Kawano and Yanai, 2013	Color and SURF	SVM		-	81.6%	[27]
Farinella et al., 2014	Bag of textons	SVM	PFID	31.3%	-	[24]
Yang et al., 2010	Pairwise local features	SVM		78.0%	-	[28]
He et al., 2014	DCD, MDSIFT, SCD, SIFT	KNN	TADA	64.5%	-	[29]
Zhu et al., 2015	Color, texture and SIFT	KNN		70.0%	-	[10]
Matsuda et al., 2012	SIFT, HOG, Gabor texture and color	MKL-SVM	UEC-Food-100	21.0%	45.0%	[9]
Liu et al., 2016	Extended HOG and Color	Fisher Vector		59.6%	82.9%	[36]
Kawano and Yanai, 2014	Color and HOG	Fisher Vector		65.3%	-	[39]
Yanai and Kawano, 2015	Color and HOG	Fisher Vector		65.3%	86.7%	[35]
Kawano and Yanai, 2014	Fisher Vector, HOG and color	One x rest Linear classifier	UEC-Food-256	50.1%	74.4%	[38]
Yanai et al., 2015	Color and HOG	Fisher Vector		52.9%	75.5%	[35]
Deep Learning Methods						
Study	Approach		Dataset	Top1	Top5	Reference
Anthimopoulos et al., 2014	ANNnh		Diabetes	75.0%	-	[31]
Bossard et al., 2014	Food-101		Food-101	56.4%	-	[18]
Yanai and Kawano, 2015	DCNN-Food			70.4%	-	[35]

Traditional Methods						
Study	Approach		Database	Performance **		Reference
	Features	Classifier		Top1 Acc.	Top5 Acc.	
Liu et al., 2016	DeepFood			77.4%	93.7%	[36]
Meyers, 2015	GoogleLeNet			79.0%	-	[11]
Hassannejad et al., 2016	Inception v3			88.3%	96.9%	[32]
Meyers, 2015	GoogleLeNet		Food201 segmented	76.0%	-	[11]
			Menu-Match	81.4% *	-	
Christodoulidis et al., 2015	Patch-wise CNN		Own database	84.90%	-	[34]
Pouladzadeh et al., 2016	Graph cut+Deep Neural Network			99.0%	-	[40]
Kawano and Yanai, 2014	OverFeat+Fisher Vector		UEC-Food-100	72.3%	92.0%	[39]
Liu et al., 2016	DeepFood			76.3%	94.6%	[36]
Yanai and Kawano, 2015	DCNN-Food			78.8%	95.2%	[35]
Hassannejad et al., 2016	Inception v3			81.5%	97.3%	[32]
Chen and Ngo, 2016	Arch-D			82.1%	97.3%	[23]
Liu et al., 2016	DeepFood			54.7%	81.5%	[36]
Yanai and Kawano, 2015	DCNN-Food		UEC-Food-256	67.6%	89.0%	[35]
Hassannejad et al., 2016	Inception v3			76.2%	92.6%	[32]
Ciocca et al., 2016	VGG		UNIMINB2016	78.3%	-	[8]
Chen and Ngo, 2016	Arch-D		VIREO	82.1%	95.9%	[23]

\* Represents the mean average precision

\*\* Top 1 and/or Top 5 indicate that the performance of the classification model was evaluated based on the first assigned class with the highest probability and/or the top 5 classes among the prediction for each given food item, respectively



**Table 4:**

Methods for food volume and calorie estimation.

Study (Year)	Approach	Performance	Reference
Noronha et al., 2011	Via crowdsourcing (e.g. users from Amazon Mechanical Turk)	Better performance than other commercial app using crowdsourcing but overall it is error prone since users estimate food portion by just looking at the picture	[42]
Chen, 2012	Use depth camera to acquire color and depth	Preliminary result showing some limitations when estimating quantity of cooked rice and water	[22]
Villalobos et al., 2012	Use Top+side view pictures with user's finger as reference	Results change due to illumination conditions and image angle; standard error is in an acceptable range	[44]
Beijbom et al., 2015	Use menu items from nearby restaurants	Food calorie is from pre-defined restaurant's menu	[20]
Meyers, 2015	3D volume estimation by capturing images with a depth camera and reconstructing image using Convolutional Neural Network and RANSAC	Using toy food; the CNN volume predictor is accurate for most of the meals; no calorie estimation outside a controlled environment.	[11]
Woo et al., 2010	Use a checkerboard as reference for camera calibration and 3D reconstruction	Mean volume error of 5.68% on a test of sever food items	[43]