MANGO - Mobile Augmented Reality with Functional Eating Guidance and Food Awareness

Georg Waltner[1(✉)], Michael Schwarz[2], Stefan Ladstätter[2], Anna Weber[2], Patrick Luley[2], Horst Bischof[1], Meinrad Lindschinger[3], Irene Schmid[3], and Lucas Paletta[2]

[1] Graz University of Technology, Graz, Austria
{waltner,bischof}@icg.tugraz.at
[2] JOANNEUM RESEARCH Forschungsgesellschaft MbH, Graz, Austria
{michael.schwarz,stefan.ladstaetter,anna.weber,
patrick.luley,lucas.paletta}@joanneum.at
[3] Institute for Nutritional and Metabolic Diseases, Schwarzl Outpatient Clinic, Lassnitzhöhe, Austria
office@lindschinger.at

Abstract. The prevention of cardiovascular diseases becomes more and more important, as malnutrition accompanies today's fast moving society. While most people know the importance of adequate nutrition, information on advantageous food is often not at hand, such as in daily activities. Decision making on individual dietary management is closely linked to the food shopping decision. Since food shopping often requires fast decision making, due to stressful and crowded situations, the user needs a meaningful assistance, with clear and rapidly available associations from food items to dietary recommendations. This paper presents first results of the Austrian project (MANGO) which develops mobile assistance for instant, situated information access via Augmented Reality (AR) functionality to support the user during everyday grocery shopping. Within a modern diet - the functional eating concept - the user is advised which fruits and vegetables to buy according to his individual profile. This specific oxidative stress profile is created through a short in-app survey. Using a built-in image recognition system, the application automatically classifies video captured food using machine learning and computer vision methodology, such as Random Forests classification and multiple color feature spaces. The user can decide to display additional nutrition information along with alternative proposals. We demonstrate, that the application is able to recognize food classes in real-time, under real world shopping conditions, and associates dietary recommendations using situated AR assistance.

Keywords: Mobile application · Video based food recognition · Augmented reality · Nutrition recommender · Functional eating concept

1 Introduction

There is a scientific consensus that modern eating habits - including excessive use of salt, fat and sugar - are one of the key factors resulting in cardiovascular

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 425–432, 2015.
DOI: 10.1007/978-3-319-23222-5_52

Fig. 1. Innovative interface functionality of the mobile application: The user applies the app similar to a "functional eating lens", moving the smart-phone in video camera modus over food appearances (left) and receives responses about the degree of recommendation from a functional eating perspective (green bars) on the recognized food items, with alternatives fitting to his nutrition plan. She can access more details and health recommendations associated with the selected food item.

diseases (CVD) [1]. According to the WHO [2], more than half of all deaths across the European Region are caused by CVD. National and European institutions expend much effort into long term prevention and information strategies, including distribution of booklets and other literature. While the literature on nutrition is steadily growing and information is readily available on the Internet, it is not practicable to carry books around or search online for specific nutrition facts when going grocery shopping. In the New Media Age, almost every person is equipped with devices such as mobile phones or tablets. Surprisingly, only very few applications are available that were developed with scientific background and can be easily utilized from persons without prior knowledge. We present early results with the aim to provide the nutrition aware user a mobile augmented reality technology that enables situated dietary information assistance and through this simplifies food choices with global aim to improves the user's overall health and well-being. We resort to the functional eating diet, devised by an experienced physician and based on investigation of a large dataset (16k entries). With focus on oxidative stress it defines pillars that stand for different stress types and are connected to nutrition recommendations; following these recommendations enables a person to alleviate stress and reduce the risk of CVD. By integrating a real-time image recognition system into the app, we facilitate the information retrieval which provides added benefit to the user (Fig. 1). This system is evaluated on a novel grocery dataset (containing fruit and vegetable classes) which had to be created from scratch, as existing food datasets concentrate on restaurant dishes only.

2 Related Work

A variety of dietary apps are commercially available, which rely on manual input of food and food volume (e.g. taking pictures or using food databases). These apps no longer focus on solely counting calorie intake, but provide tailored diet plans (e.g. Diet Point Weight Loss) or detailed nutritional information (e.g. LoseIt). Popular fitness apps like Endomondo and Runkeeper (MyFitnessPal) integrate data from wireless scales or activity trackers (Fitbit, Withings,...). Community support ensures a steady growth of food databases (e.g. MyFitnessPal database with over 5 million foods), providing motivation by competing with other users or receiving feedback. Some apps assist with food logging or a integrated barcode scanner; e.g. ShopWell is a shopping assistant rating scanned groceries according to a personalized profile (health conditions, food allergies, athletic training) and provides health recommendations.

For image-based food recognition a large amount of research prototypes exist (e.g. [3]), however, hardly any commercial applications are available. A promising work was introduced by [4], capable of identifying food and portion size for calorie count estimation. The application also incorporates contextual features (restaurant locations, user profiles) for refinement and result augmentation. Besides mobile food recognition approaches, more sophisticated methods mostly focus on distinguishing dishes or ingredients. [5] classify images in food and non-food categories under user interaction utilization. Multiple Kernel Learning (MKL) is examined by [6], [7] use pairwise features and [8] explore the use of a bag-of-texton framework. [9] use depth information to classify and quantify Chinese food categories. A candidate region detection process is used by [10], they classify food with a Support Vector Machine (SVM). [11] learn a Bag-of-Features (BoF) model on HSV-SIFT and color moment features to estimate carbohydrate content for diabetic patients. [12] present a combination of Random Forests (RF) and SVM to mine typical image components which are discriminative for 101 food categories collected from the web. Deep convolutional features and a SVM classifier are used by [13]. In the context of fruit and vegetable recognition, [14] present a system for facilitating the supermarket checkout process. [15] use a feed-forward neural network for fruit recognition. An overview of commercial vision systems can be found in [16], [17] and [18]. Similar to our method, [19] present a mobile cooking recipe recommendation system employing object recognition for food ingredients. Their approach is based on a BoF (SURF and color histograms) using a linear SVM classifier. In contrast to our method, most afore-mentioned approaches are intractable to handle on mobile devices due to the complexity of the recognition systems. Also the images are often manually selected or acquired in a controlled laboratory environment, while our data is collected in supermarkets from individuals without background on computer vision. Additionally, we employ recognition in a temporal context, as our inputs are videos and not images. While incorporating a barcode reader and/or Optical Character Recognition (OCR) could improve recognition, we want to create a general system independent of store specific barcodes, price tags or food packages. We focus mainly on unpackaged, loose grocery items.

3 Functional Eating Dietary Concept

Functional eating is a modern diet, developed based on medical expertise particularly concerning demand oriented nutrition. On a dataset including over 16000 entries, [20,21] show that the measurement of oxidative stress is one of the main factors which can be alleviated by a an optimal food combination. Combining food in a certain way further results in health benefits and improved performance, accompanied by a better taste experience. Often only few fruit types are able to cover the demand for individual vitamins and minerals, therefore a correct food combination is directly connected to the general well-being and health of people, as well as an increased mental and physical performance. Age- and gender-related diversities and differing stress profiles (with respect to work, family or sport) lead to person-specific nutrition profiles. As proposed by [22], functional eating addresses oxidative stress reduction by presenting seven pillars for different nutrition requirements, they are depicted in Fig. 2. These pillars are incorporated into a mobile application that provides feedback on the user's food choices during grocery shopping. The user gets recommendations triggered by his individual profile and is presented detailed nutrition information about food items.

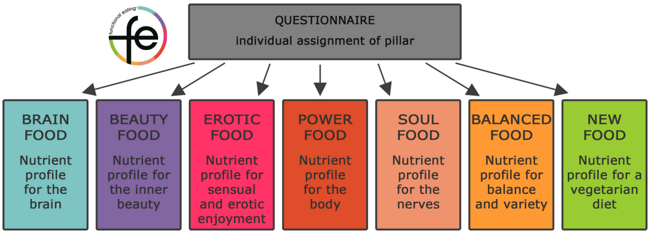

Fig. 2. Functional eating offers seven pillars corresponding to specific nutrition guidelines. After questionnaire completion, a person is assigned to one of the tailored diet pillars.

4 Mobile Augmented Reality and Information System

The MANGO project aims at the development of a mobile application with situated nutrition assistance for dietary management on the basis of an augmented reality based recommender component. The food information and recommender assistant provides an intuitive interface with video based food recognition. The nutritional advice in the frame of functional eating is tailored to fit the user's profile which has been assessed by a dietary questionnaire in advance. After scanning a food item, the food recognition result is shown on the mobile display, together with alternatives for nutrition, associated with a ranking bar indicating how well the food matches the dietary user profile. By clicking on a result, detailed information is shown, including macronutrients with corresponding health claims as well as further food recommendations matching the user's profile (Fig. 1). Complementing the image-based food recognition, the app allows to browse or search through a rich food database specially created for the application in the global context of functional eating. Users might also read an introduction about the

(a) (b) (c) (d) (e) (f)

Fig. 3. (a) On the welcome screen, the user chooses between camera mode, a food database and functional eating information. (b) Food database with information about ingredients and health claims. (c,d) Information screens about functional eating and the beauty food pillar. (e,f) The user fills in a questionnaire to determine his nutrition profile and is assigned to the most suitable pillar.

underlying functional eating nutrition concept. See Fig. 3 for app details and Fig. 1 for a usage example.

4.1 Food Recognition Methodology

For video (image) classification we use a Random Forest, a classifier that gained popularity after introduction in [23]. RF have multiple advantages over other classifiers like Support Vector Machines or Neural Networks (NN). They are capable of handling large training datasets, are inherently suited for multiclass problems, avoid overfitting due to random feature selection and are comparably fast to train and evaluate. A random forest is an ensemble of decision trees, which are trained separately by random feature selection and choosing the optimal node splits from this feature pool. Arriving samples at every forest leaf are stored as class probability histograms; during testing the results of all trees are averaged and the class with the maximum average probability is selected as classification result. We use the discriminative color descriptor [24] and a combination of color histograms, namely RGB, HSV and LAB and concatenate the features to a single vector. We have also explored different shape and texture descriptors like SURF [25], SIFT [26], HOG [27], PHOG [28], wavelet moments or Gabor wavelets. Since the accuracy did not improve significantly and the cost in terms of computing time was too high for real time performance on mobile devices, we removed them from the framework. The features are calculated on image patches in a sliding window fashion. After patch classification, rather than taking the cumulative maximum of the predicted patch classes, we fuse the results by calculating the mean and median over all patch histograms and take the averaged maximum of both as classification result. This gives robustness to outliers and minors their impact on the final result.

5 Experimental Results

For evaluation we recorded 1719 videos from different supermarkets and generated a dataset containing 23 classes (e.g. tomatoes) with 98 subclasses (e.g. tomatoes on the vine, mixed tomatoes, beef tomatoes, Kumato tomatoes). The videos were recorded with five different standard android smartphones (Full HD). We evaluate our experiments on the entire dataset, as well as on a smaller sub-dataset of 12 classes for the mobile application prototype. These classes were chosen from seasonal available food groups at that time, to allow for an in-market user study. From the videos we extracted 150 images per class, samples can be seen in Fig. 4. We perform experiments on a desktop PC to evaluate our

Fig. 4. Datasets: Sample images for each of the 35 food classes and the 12 classes used in the prototype application. From top left to bottom right: cherries, apricots, strawberries, blackberries, blueberries, chanterelles, champignons, tomatoes on the vine, green salad, pears, broccoli, cauliflower, cabbage, grapes, bananas, herbs, horseradish, plums, damsons, raspberries, red currants, black currants, white currants, brown mushrooms, red apples, green apples, mixed tomatoes, beef tomatoes, Kumato tomatoes, iceberg salad, Lollo rosso salad, yellow peppers, red peppers, green peppers, mixed peppers.

approach with a greater amount of classes using the entire dataset. By dividing some of the 23 classes according to intra-class differences (e.g. green/red apples), we obtain 35 classes, which are evaluated by the RF with a 50%/50% split, using previously discussed features. The mobile classifier (app) is trained and evaluated with a 80%/20% split on the sub-dataset of 12 classes. In addition to the top vote, we also calculate bullseye scores for top 2 and top 3 votes for the classifier, as in the app's food scanning mode three proposals are displayed to the user. We achieve a classification rate of 80.30% for predicting the correct food category with the maximum confidence vote, and subsequently a rate of 92.28% for the top 2 and 97.20% for the top 3 class classifications. Our first prototype runs at 5fps. Table 1 shows a comparison of our two approaches. On one hand, the mobile prototype results evaluated on the 12 class sub-dataset (MANGO-12) on the other hand the evaluation on the entire dataset (MANGO-35). Our method could not be evaluated on the most similar mobile approach of [19], as their dataset is not publicly available.

Table 1. Evaluation on the MANGO-12 and MANGO-35 datasets on a mobile phone and a desktop PC respectively. Top 1 to top 3 bullseye scores are displayed.

Dataset	Classes	Acquisition	Classifier	Top 1	Top 2	Top 3
MANGO-12	12	mobile phone	mobile phone	80.30%	92.28%	97.20%
MANGO-35	35	mobile phone	desktop PC	67.25%	85.97%	91.97%

6 Conclusion

We have presented an innovative mobile assistant solution, that supports the user in his grocery shopping and situated dietary management decisions . Within the app we incorporated the functional eating diet as scientific basis and included an early, promising prototype of an automated video based food recognition component. Future developments will include more food classes and enhanced adaptive classifiers, which are a main subject of ongoing research. As there exist only few public food datasets (such as [10,12,29]) and these contain dishes in contrast to grocery items, we plan to publish our fruit and vegetable dataset in the future along with fine and coarse category labels.

Acknowledgments. This work was supported by the Austrian Research Promotion Agency (FFG) under the project Mobile Augmented Reality for Nutrition Guidance and Food Awareness (836488).

References

1. Waxman, A., Norum, K.R.: Why a global strategy on diet, physical activity and health? The growing burden of non-communicable diseases. Public Health Nutrition **7**, 381–383 (2004)
2. World Health Organization: European Action Plan for Food and Nutrition Policy, pp. 2007–2012 (2008)
3. Oliveira, L., Costa, V., Neves, G., Oliveira, T., Jorge, E., Lizarraga, M.: A mobile, lightweight, poll-based food identification system. Pattern Recognition **47**(5), 1941–1952 (2014)
4. Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., Sawhney, H.: “Snap-n-Eat” Food Recognition and Nutrition Estimation on a Smartphone. DST (2015)
5. Maruyama, Y., de Silva, G.C., Yamasaki, T., Aizawa, K.: Personalization of food image analysis. In: VSMM, pp. 75–78 (2010)
6. Hoashi, H., Joutou, T., Yanai, K.: Image recognition of 85 food categories by feature fusion. In: ISM, pp. 296–301 (2010)
7. Yang, S., Chen, M., Pomerleau, D., Sukthankar, R.: Food recognition using statistics of pairwise local features. In: CVPR, pp. 2249–2256 (2010)
8. Farinella, G.M., Moltisanti, M., Battiato, S.: Classifying food images represented as bag of textons. In: ICIP, pp. 5212–5216 (2014)
9. Chen, M.Y., Yang, Y.H., Ho, C.J., Wang, S.H., Liu, S.M., Chang, E., Yeh, C.H., Ouhyoung, M.: Automatic Chinese Food Identification and Quantity Estimation. In: SIGGRAPH, pp. 29:1–29:4 (2012)

10. Matsuda, Y., Hoashi, H., Yanai, K.: Recognition of multiple-food images by detecting candidate regions. In: ICME, pp. 25–30 (2012)
11. 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. JBHI **18**(4), 1261–1271 (2014)
12. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101 – mining discriminative components with random forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VI. LNCS, vol. 8694, pp. 446–461. Springer, Heidelberg (2014)
13. Kawano, Y., Yanai, K.: Food image recognition with deep convolutional features. In: UbiComp Adjunct, pp. 589–593 (2014)
14. Bolle, R.M., Connell, J.H., Haas, N., Mohan, R., Taubin, G.: Veggievision: a produce recognition system. In: WACV, pp. 244–251 (1996)
15. Zhang, Y., Wang, S., Ji, G., Phillips, P.: Fruit classification using computer vision and feedforward neural network. Journal of Food Engineering **143**, 167–177 (2014)
16. Jiménez, A.R., Jain, A.K., Ceres, R., Pons, J.L.: Automatic fruit recognition: a survey and new results using Range/Attenuation images. Pattern Recognition **32**(10), 1719–1736 (1999)
17. Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., Liu, C.: Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. Food Research International **62**, 326–343 (2014)
18. Costa, C., Antonucci, F., Pallottino, F., Aguzzi, J., Sun, D.W., Menesatti, P.: Shape Analysis of Agricultural Products: A Review of Recent Research Advances and Potential Application to Computer Vision. FABT **4**(5), 673–692 (2011)
19. Maruyama, T., Kawano, Y., Yanai, K.: Real-time mobile recipe recommendation system using food ingredient recognition. In: IMMPD Workshop, pp. 27–34 (2012)
20. Lindschinger, M., Nadlinger, K., Adelwöhrer, N., Holweg, K., Wögerbauer, M., Birkmayer, J., Smolle, K.H., Wonisch, W.: Oxidative stress: potential of distinct peroxide determination systems. CCLM **42**(8), 907–914 (2004)
21. Wonisch, W., Falk, A., Sundl, I., Winklhofer-Roob, B., Lindschinger, M.: Oxidative stress increases continuously with bmi and age with unfavourable profiles in males. Aging Male **15**(3), 159–165 (2012)
22. Karalus, B., Lindschinger, M.: Eat yourself beautiful, smart and sexy with functional eating (in German). Riva Verlag, Munich (2008)
23. Breiman, L.: Random Forests. Machine Learning **45**(1), 5–32 (2001)
24. Khan, R., van de Weijer, J., Khan, F.S., Muselet, D., Ducottet, C., Barat, C.: Discriminative color descriptors. In: CVPR, pp. 2866–2873 (2013)
25. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
26. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. IJCV **60**(2), 91–110 (2004)
27. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: CVPR, vol. 1, pp. 886–893 (2005)
28. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: CIVR, New York, NY, USA, pp. 401–408 (2007)
29. Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R.: PFID: pittsburgh fast-food image dataset. In: ICIP, pp. 289–292 (2009)