

Personalized Dietary Self-Management Using Mobile Vision-Based Assistance

Georg Waltner¹(✉), Michael Schwarz², Stefan Ladstätter², Anna Weber², Patrick Luley², Meinrad Lindschinger³, Irene Schmid³, Walter Scheitz⁴, Horst Bischof¹, and Lucas Paletta²

¹ Graz University of Technology, Graz, Austria
waltner@icg.tugraz.at

² Joanneum Research Forschungsgesellschaft mbH, Graz, Austria

³ Institute for Nutritional and Metabolic Diseases, Lassnitzhöhe, Austria

⁴ FH Joanneum University of Applied Sciences, Graz, Austria

Abstract. Daily appropriate decision making on nutrition requires application of knowledge where it matters, and being adjusted to the individual requirements. We present a highly personalized mobile application that assists the user in appropriate food choices during grocery shopping, while simultaneously incorporating a personalized dietary recommender system. The application can be used in video based augmented reality mode, where a computer vision algorithm recognizes presented food items and thus replaces tedious search within the food database. The recognition system employs a shallow Convolutional Neural Network (CNN) based classifier running at 10 fps. An innovative user study demonstrates the high usability and user experience of the application. The vision classifier is evaluated on a newly introduced reference image database containing 81 grocery foods (vegetables, fruits).

1 Introduction

Today's population lives in a fast-paced society with challenging work environments, a multitude of free time activities available and increasingly less knowledge about food origins and nutritional value. Intelligent dietary self-management systems can save time, improve personal nutrition and such lead to a healthier living and decreased stress. In recent programs like the *Food Scanner Prize* of the European Commission (EC), institutions and authorities expend much effort into developing systems that reduce food-related problems.

Such systems should comply with four general demands. First, it should be a portable solution packed into a small mobile device. Second, it must be simple to use without prior knowledge in nutrition concerns and with basic computer skills. Third, the applied system should work fast and reliable. Finally, it should be able to provide valuable feedback to users regarding their health and lifestyle, resulting in better decision making.

We propose a system that aims to comprise all requirements and is intended to be used by nutrition-aware persons. Our system is developed as a portable,

easy to use system running on a smart phone. It enables situated dietary information assistance and simplifies food choices with the aim to improve the user’s overall health and well-being. Our proposed system has two core components (Fig. 1). First, we have implemented a mobile application that integrates personalized dietary concerns into a recommendation system used during grocery shopping. This diet has been developed by medical experts in the field of nutrition and metabolism [18]. The user can take a survey for a customized grocery basket and additionally enter information about her personal condition, which is used for calculating personal energy expenditure. Second, we have developed a fast lightweight computer vision component that enables the user to retrieve information from the food database by pointing the device to grocery items for automatic recognition, instead of tediously entering information by hand. The image recognition system runs with high accuracy on a big set of different grocery food classes. It facilitates information retrieval, providing added benefit to the user.

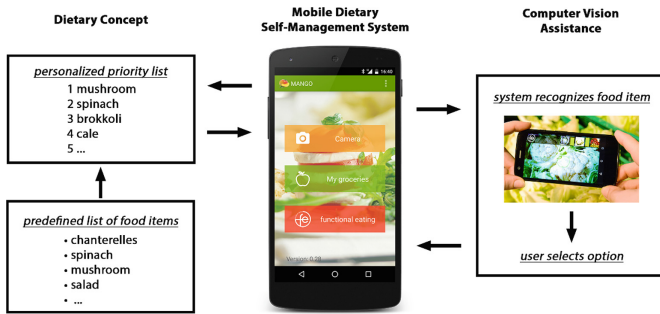


Fig. 1. System overview. The dietary self-management system consists of two core elements: a personalized list of recommended groceries and computer vision based assistance for information retrieval.

2 Related Work

Dietary Mobile Applications. Mobile health and wellness is a rapidly expanding market, with daily emergence of innovative dietary management apps. For example, *LoseIt*¹ aids the user in losing weight by setting daily calorie limits and monitoring food intake. It also features a recognition system which is coupled to a database of dishes where the user must select the appropriate one. In contrast, our system is designed to aid already during the food selection process in grocery stores and targets a more general audience that wants to better its eating behaviour. *ShopWell*² rates scanned foods and provides appropriate recommendations according to a personalized profile. The scanning works

¹ <https://www.loseit.com>.
² <http://www.shopwell.com>.

for barcodes only, which is also available in our application besides automated video recognition. Several EC research programs fund the investigation of dietary management, such as, for the care of elderly. *CordonGris*³ manages relevant data for a healthy diet recommendation, coming from different sources: activity sensors, food composition tables, retailers' or service providers' information. *HELICOPTER*⁴ exploits ambient-assisted living techniques and provides older adults and their informal caregivers with support, motivation and guidance in pursuing a healthy and safe lifestyle, including decision making on nutrition in grocery shopping. *ChefMySelf*⁵ is a customizable open and extensible eco-system built around an automatic cooking solution to support elderly in preparing healthy meals.

Food Recognition Systems. Most research in computer vision based methods targets the recognition of meals as well as extracting the components of plated food. First food recognition systems have already been introduced in the late 90s, "Veggie Vision" [2] eases the checkout process at supermarkets. The topic regained attention recently with published food datasets for comparison of methods, e.g., PFID [5], UNICT-FD889/1200 [7, 8], Food-101 [3], UECFOOD-100/256 [14, 21]. Until the recent rise of CNN based methods [13], that automatically learn optimal feature representations from thousands of images, researchers mostly combined handcrafted color and texture descriptors with SVM classifiers or other kernel methods. In [12], a CNN recognizes the 10 most frequent food items in the FoodLog [17] image collection and is able to distinguish food from non-food items. In [15, 16] a CNN is finetuned on 1000 food related classes from the ImageNet database [6]. Recent wider [19] and deeper [11] CNNs boosted the results at the cost of high computational requirements. Compared to our method, that is running at 10 fps on standard smart phones, most afore-mentioned classification approaches are intractable to handle on mobile devices.

Dietary Self-Management Systems. Few available applications combine dietary mobile systems with automated food recognition. In [20], a mobile recipe recommender recognizes ingredients and retrieves recipes online. A mobile application proposed in [1] supports type 1 diabetes patients in counting carbohydrates and provides insulin dose advice. "Snap-n-Eat" [27] identifies food and portion size for calorie count estimation by incorporating contextual features (restaurant locations, user profiles). None of the above aids the user as early as in the food selection stage, but only when meals are already prepared. Some also rely on an internet connection while our system is entirely running on device. Compared to [26], this work uses a lightweight CNN instead of a Random Forest classifier which allows to recognize more than double the classes. Personalization has also been lifted to a new level, comprising personalized energy expenditure

³ <http://cordongris.eu>.

⁴ <http://www.aal-europe.eu/projects/helicopter>.

⁵ <http://www.chefmyself.eu>.

calculation and target weight advice. Usability is evaluated via an innovative user study.

3 Personalized Dietary Self-Management System

The proposed system enables dietary self-management on a mobile device. It includes integrated nutrition assistance based on an augmented reality recommender component. This recommender assistant provides an intuitive interface and is supported by video based food recognition. A user specific profile is assessed by a dietary questionnaire on first use. Afterwards, upon selection of food items, either from automated video recognition or from manual user selection, tailored nutritional advice is given to the user depending on her profile.

3.1 Dietary Concept for Self-Management

The dietary concept behind the recommender system is based on a personalized dietary concept with the idea of removing stringent rules (e.g. calorie counting, physical activity demands). Instead of forcing the patient to give up all possibly bad eating habits at once, the diet slowly changes nutrition habits to lose (gain) weight. Considering that every person has its own requirements concerning nutrition and every fruit or vegetable has its own composition of micronutrients, a correct food combination is indispensable to fulfill one's individual demand of essential nutrients. The utilized diet incorporates this by defining several groups according to lifestyle, job, age, gender and the intensity of personal exercise. The groups are connected to different stress types and include nutrition recommendations, upon individual baskets of commodities that are composed based on investigation of a large medical dataset of 17,000 entries in [18]. Our mobile application incorporates these baskets and provides situated feedback on the user's food choices during grocery shopping, where decision making for mid and long-term lasting food choice is actually taking place. This aims at a lasting change in eating behaviour for increased mental and physical performance according to the *functional eating diet* [26]. The app automatically classifies presented food and upon commodity selection the user gets recommendations triggered by her individual profile and is presented detailed nutrition information. This includes micronutrients with corresponding health claims as well as further food recommendations matching the user profile.

3.2 Personalization

Personalization of the self-management system is based on two main factors. First, the assignment of the user to a certain nutrition group (see [26]) and, second, customized energy expenditure calculation. The questionnaire for user to group assignment consists of multiple rating scale questions from whom the user type can be calculated, e.g. "Do you often feel stressed?" with possible answers between "1, not at all" to "4, very often". We will now describe how the

personalized energy expenditure is calculated, as it is influenced by the height, age, body type and physical activity level (PAL) of a person.

To account for different body types (slim, normal, muscular) and gender differences, the body structure of a person is incorporated through a weighting term δ , that ranges from 0.945 (slim) to 1.055 (muscular) for men and 0.900 to 1.000 for women respectively. A personalized target weight w_p is calculated by multiplication of the body weight w with δ : $w_p = w * \delta$. The energy demand E as defined by [10] is adopted to the newly calculated weight with $E^{\sigma} = 66.473 + (5.003 * l * 100) + (13.752 * w_p) - (6.755 * \alpha)$ for men and $E^{\varphi} = 655.096 + (1.85 * l * 100) + (9.563 * w_p) - (4.676 * \alpha)$ for women, where l is the height and α is the age of the person. Finally, PAL are considered through a factor γ_{PAL} . They reflect the energy demand in dependence of physical activity and are such very suitable for personalized energy expenditure calculation. PAL factors range from $\gamma_{PAL} = 1.2$ for elderly people without any physical activity to $\gamma_{PAL} = 3.3$ for construction workers spending 20+ hours on sport. The final personalized daily energy expenditure is calculated as $E_p = E * \gamma_{PAL}$.

3.3 Mobile Recognition System

We improve the usability of the recommender system and automatically classify food items with Convolutional Neural Networks (CNN). The user taps on an item, confirms the classification result and is displayed the desired information instead of cumbersome manual search. Our motivation is to get a fast, scalable classifier and we implement a shallow CNN network, running within our Android based mobile application at 10 fps. We design our CNN to have minimum complexity while performing at good accuracy.

4 Experimental Results

We evaluate our system with respect to usability through a user study and measure the performance of the recognition system on a novel grocery database.

4.1 Usability

For the purpose of user-centered optimisation of the novel computer vision based interface design, an innovative interaction and usability analysis was performed with 16 persons, $M = 26.3$ years of age. We used eye tracking to evaluate the automated nutrition information feedback interface component and evaluated the user experience in the frame of using the complete app. Eye tracking as a method to evaluate novel interface designs has been established, for example, using fixation durations on objects in the user interface: depending on the context, high numbers of fixations indicate less efficient search strategies, long fixation durations indicate difficulties of the user with the perception of the display [9]. Test persons were equipped with SMITM eye tracking glasses, the viewing behavior was video captured, and fixations on the display were localised

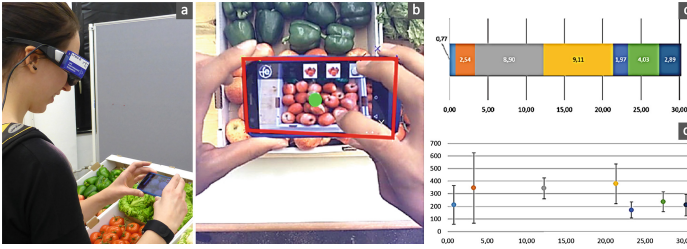


Fig. 2. Innovative interaction and usability analysis: (a) mobile eye tracking glasses (b) automated gaze localisation and analysis of seven stages of interaction from (c) stage duration and (d) corresponding fixation analysis.

[23]. From the investigation of the *Seven Stages of (Inter-)Action* [22] and corresponding fixation analysis interaction design was updated and optimised towards a SUS usability score [4] of 80% and user experience evaluation (UEQ [25]) of 72% ($\pm 5\%$, 90% confidence interval; [24]) which represent high scores considering the early stage of development. See Fig. 2 for illustrations.

4.2 Recognition System

We evaluate our method on a newly recorded dataset, which we term *FruitVeg-81*⁶. The database contains 15630 images of 81 raw fruit and vegetable classes. The proposed CNN processes images with a size of 56×56 pixels and consists of three convolutional layers: the first two are of size $5 \times 5 \times 32$, the third layer is $5 \times 5 \times 64$. We apply pooling with size 3 and stride 2 after each layer. After the third convolutional layer we add a 1024-dimensional fully connected layer with dropout and a soft-max classification layer with 81 units for food classification and 82 units when integrating a garbage class. We subtract the training set mean and train the network in minibatches of size 128, the training set is shuffled in the beginning of the training procedure. As it is unlikely to reach 100% accuracy in practical use, we give the users several choices for selecting the correct food item and reflect this in our experiments by reporting the *top-1* to *top-5*.

Baseline. As Baseline, we evaluate all models on the 81 classes using leave-one-out cross-validation. We augment the training data with mirroring, cropping, rotating and color shifting. During test time we use mirroring and random crops.

Non-food Class. For our real-world application it is important to reduce false positives, e.g. on food items missing in the visual database. When the application recognizes non food items, appropriate feedback is displayed on the screen. We

⁶ <https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/fruitveg81/>.

extract around 200 random images from 500 non-food categories of the ImageNet Challenge [6] and use those to add another category to the CNN. Due to the high variance within this non-food class, we choose the amount of images for training to be 10 times the average number of images per food category. For testing, we use the average number of per class test images.

Results are listed in Table 1, it can be seen that the image quality of the mobile phones differs. On average a top-5 accuracy of around 90% shows the good performance of the trained network. As we add one more class to the system the accuracy decreases, however the resulting decrease of the *top-1* mean accuracy is stronger than expected. This is presumably due to the very heterogeneous structure and different domain of the non-food samples, which is hard to model with the limited number of parameters. On the other hand the *top-5* accuracy is stable, which is a desired behavior for our application.

Table 1. Results for baseline and integration of a non-food class. The mean *top-k* accuracy ranges from 69.77% to 90.19% for the baseline and from 60.47% to 90.41% for non-food integration (best *top-1* accuracy is 76.14% and 71.74%). With non-food integration the *top-1* mean accuracy drops by roughly 9%, while the *top-5* mean accuracy remains the same.

Model	Baseline					Non-food				
	Top-1	Top-2	Top-3	Top-4	Top-5	Top-1	Top-2	Top-3	Top-4	Top-5
Samsung Galaxy S3	72.99	84.34	88.93	92.27	94.58	64.45	80.72	87.55	91.15	93.39
Samsung Galaxy S5	76.14	86.66	90.82	92.46	93.82	71.74	84.30	89.11	92.69	94.52
HTC	71.99	84.06	87.86	89.77	91.03	60.17	77.89	84.49	89.00	91.77
HTC One	65.28	76.95	82.84	85.65	88.40	52.30	71.32	80.41	85.50	88.14
Motorola G	62.43	73.72	78.22	81.00	82.94	53.70	68.54	75.17	79.98	84.21
Avg	69.77	81.15	85.72	88.23	90.16	60.47	76.53	83.35	87.66	90.41

5 Conclusion

We have presented a innovative mobile application with a recommender engine and a fast recognition system running at 10 fps as core elements. The recommender engine supports users in decision making during grocery shopping and helps to improve health conditions backed on scientific findings. The recognition system is robustly recognizing food and non-food items. Along with this publication we make our grocery dataset *FruitVeg-81* available for the public, with the intention to be used by scientific researchers around the globe for improvement of their nutrition related computer systems.

Acknowledgments. This work was supported by the Austrian Research Promotion Agency (FFG) under the project MANGO (Mobile Augmented Reality for Nutrition Guidance and Food Awareness), No. 836488.

References

1. Anthimopoulos, M., Dehais, J., Diem, P., Mougiakakou, S.: Segmentation and recognition of multi-food meal images for carbohydrate counting. In: Proceedings of the BIBE, pp. 1–4 (2013)
2. Bolle, R.M., Connell, J.H., Haas, N., Mohan, R., Taubin, G.: VeggieVision: a produce recognition system. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 244–251 (1996)
3. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101 - mining discriminative components with random forests. In: Proceedings of the ECCV, pp. 446–461 (2014)
4. Brooke, J.: SUS-A quick and dirty usability scale. *Usability Eval. Ind.* **189**(194), 4–7 (1996)
5. Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R., Yang, J.: PFID: Pittsburgh fast-food image dataset. In: Proceedings of the ICIP, pp. 289–292 (2009)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: Proceedings of the CVPR (2009)
7. Farinella, G.M., Allegra, D., Stanco, F.: A Benchmark dataset to study the representation of food images. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8927, pp. 584–599. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16199-0_41
8. Farinella, G.M., Moltisanti, M., Battiato, S.: Classifying food images represented as bag of textons. *Comput. Biol. Med.* **77**(C), 23–39 (2016)
9. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. *Int. J. Ind. Ergon.* **24**(6), 631–645 (1999)
10. Harris, J.A., Benedict, F.G.: A biometric study of human basal metabolism. *PNAS* **4**(12), 370–373 (1918)
11. Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., Cagnoni, S.: Food image recognition using very deep convolutional networks. In: Proceedings of the MADiMa, pp. 41–49 (2016)
12. Kagaya, H., Aizawa, K., Ogawa, M.: Food detection and recognition using convolutional neural network. In: Proceedings of the ACM MM, pp. 1085–1088 (2014)
13. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: Proceedings of the CVPR, pp. 1725–1732 (2014)
14. Kawano, Y., Yanai, K.: Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8927, pp. 3–17. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16199-0_1
15. Kawano, Y., Yanai, K.: Food image recognition using deep convolutional features pre-trained with food-related categories. In: Proceedings of the MBDA Workshop (2014)
16. Kawano, Y., Yanai, K.: Food image recognition with deep convolutional features. In: Proceedings of the UbiComp Adjunct, pp. 589–593 (2014)
17. Kitamura, K., Yamasaki, T., Aizawa, K.: Food log by analyzing food images. In: Proceedings of the ICM, pp. 999–1000. ACM (2008)
18. 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. *Clin. Chem. Lab. Med.* **42**(8), 907–914 (2004)
19. Martinel, N., Foresti, G.L., Micheloni, C.: Wide-Slice Residual Networks for Food Recognition, arXiv preprint [arXiv:1612.06543](https://arxiv.org/abs/1612.06543) (2016)

20. Maruyama, T., Kawano, Y., Yanai, K.: Real-time mobile recipe recommendation system using food ingredient recognition. In: Proceedings of the IMMPD, pp. 27–34 (2012)
21. Matsuda, Y., Hoashi, H., Yanai, K.: Recognition Of multiple-food images by detecting candidate regions. In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 25–30 (2012)
22. Norman, D.A., Draper, S.W.: User Centered System Design: New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., Hillsdale, NJ 3 (1986)
23. Paletta, L., Neuschmied, H., Schwarz, M., Lodron, G., Pszeida, M., Ladstätter, S., Luley, P.: Smartphone eye tracking toolbox: accurate gaze recovery on mobile displays. In: Proceeding of the Symposium on Eye Tracking Research and Applications, pp. 367–368. ACM (2014)
24. Rexeis, V.: Usability Benchmark und Aktivitäts-Analyse mit Eye Tracking von Mobile Augmented Reality unterstützten Ernährungsempfehlungen. Master's thesis, Graz University of Technology (2015)
25. Schrepp, M., Olschner, S., Schubert, U.: User Experience Questionnaire Benchmark: Praxiserfahrungen zum Einsatz im Business-Umfeld (2013)
26. Waltner, G., Schwarz, M., Ladstätter, S., Weber, A., Luley, P., Bischof, H., Lindschinger, M., Schmid, I., Paletta, L.: MANGO - mobile augmented reality with functional eating guidance and food awareness. In: Murino, V., Puppo, E., Sona, D., Cristani, M., Sansone, C. (eds.) ICIAP 2015. LNCS, vol. 9281, pp. 425–432. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23222-5_52
27. Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., Sawhney, H.: Snap-n-Eat: food recognition and nutrition estimation on a smartphone. *JDST* **9**(3), 525–533 (2015)