DOI: 10.58895/ksp/1000128686-4 erschienen in:
OCM 2021 - 5th International Conference on Optical Characterization of Materials, March 17th – 18th, 2021, Karlsruhe, Germany : Conference Proceedings
DOI: 10.58895/ksp/1000128686 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000128686/

Towards the universal assessment of dietary intake using spectral imaging solutions

Yannick Weesepoel[1], Freek Daniels[2], Martin Alewijn[1], Mireille Baart[3], Judith Müller-Maatsch[1], Görkem Simsek-Senel[2], Hajo Rijgersberg[2], Jan Top[2,4], and Edith Feskens[3]

[1] Wageningen Food Safety Research, Wageningen, The Netherlands
[2] Wageningen Food and Biobased Research, Wageningen, The Netherlands
[3] Wageningen University and Research, Department of Human Nutrition and Health, Wageningen, The Netherlands
[4] Vrije Universiteit Amsterdam - Faculty of Sciences, Computer Sciences, Business, Web and Media, Amsterdam, The Netherlands

Abstract Personal(ised) sensors or "wearables" could, in the future, be applied to provide personalized nutritional advice. A challenge here is to assess dietary intake. To date, this has mainly been done with questionnaires or interviews, for example, food frequency questionnaires or 24-hour recalls. However, these methods are prone to bias due to conscious or unconscious misreports, and more objective measurement methods are desirable. By applying and combining spectral imaging techniques like hyperspectral imaging (HSI) and RGB-depth (RGBD) imaging, information on the macro-composition, identity and quantity of food consumed can be obtained. In this work, we demonstrate that HSI was effective for estimation of the fat content and layer thickness of butter on slices of bread with root mean squared errors of predictions of 4.6 (fat w/w %) and 0.056 mm respectively. Identification and volume estimation of vegetables and preparation methods were successful with RGBD imaging. Using Convolutional Neural Networks, all samples were correctly identified. For volume estimations of vegetables, R-square scores between 0.80 – 0.96 were achieved.

Keywords Non-destructive sensing, measuring and detecting, healthy behavior, consumer science

1 Introduction

Dietary intake data is important in epidemiology, for dietary interventions and for providing dietary advice. Advances in nutritional science have facilitated the shift from general population-based dietary guidelines to more personalized dietary recommendations. To provide personalized advice, an appropriate method for measuring and assessing dietary intake is required. Dietary intake is traditionally assessed with questionnaires or interviews, for example, food frequency questionnaires or 24-hour recalls. However, these methods are prone to bias due to conscious or unconscious misreports. Therefore, more objective measurement methods are desirable. Moreover, the traditional methods are rather time-consuming for both researchers and consumers.

For this reason, we aim to develop an easy-to-use measurement set-up that enables consumers to measure their dietary intake themselves, using state-of-the-art vibrational spectroscopy and imaging techniques. The analytical and machine learning challenges for this work comprise several issues:

1. Object-level recognition to identify the type(s) of food products present, to estimate the volume of those objects (before and after eating) and to obtain intake quantities from this information. For this purpose, Red Green and Blue-depth imaging (RGBD) was explored (this work).

2. Molecular-level recognition for the determination of the macro and micro composition and, to a lesser extent, identification of the product. Here a snapshot hyperspectral imaging (HSI) system in the short-wavelength infrared (SWIR) was used (this work).

3. Data fusion methods to connect the data of both imaging sensors to improve decision accuracies and versatility of the set-up.

4. Application of a functional metadata decision layer to deal for example with the within-food heterogeneity, the enormous diversity of products and dishes available [1].

In the current study, a first demonstrator is developed at technology readiness level 4. This means that the demonstrator can be operated in a controlled environment by trained persons. We demonstrate in this study the individual imaging techniques applied to the intake of butter (HSI), and vegetables (RGBD), and their respective data analysis procedures.

2 Materials and Methods

2.1 Materials

Fresh (1 – 2 days after baking) sliced white and wholegrain bread, crispbread, 44 commercially available types of butter and margarine, fresh endive (Cichorium endivia), fresh carrots (Daucus carota subsp. sativus), fresh spinach (Spinacia oleracea), frozen spinach and spinach with added cream of a single brand were purchased from a local supermarket in Wageningen, The Netherlands. All chemicals used for macro composition reference analysis were in purity grades following the specifications in the ISO standards used (Section 2.3).

2.2 Sample preparation

To test the ability of dietary intake assessment, the sensing principles were firstly assessed separately in two food cases: (i) (low-fat) margarine and butter applied to a sandwich using HSI, and (ii) raw and prepared vegetables using RGBD imaging. For the first case, standardized amounts (3, 6, 10, 15 and 30 g) of 28 butter types with varying fat concentrations (29.8% - 84.3% w/w) were applied on three different types of sandwiches with standardized sizes (wholegrain and white 10 x 10 cm and crispbread 6.0 x 8.1 cm). A randomized experimental design was made in combining the type of bread, margarine/butter type and layer thickness of the butter/margarine, resulting in 84 samples. As an alternative experiment, butter was applied (16 butter types, fat concentration 28.0% - 82.5% w/w, applied on white and wholegrain bread) using a custom-made triangular applier, resulting in a continuous layer thickness of 10 mm – $<$0.5 mm (Figure 2.1 A-B).

For the second food case, fresh endive, fresh carrots and fresh spinach were investigated in raw form and after cooking. In addi-

tion to the fresh spinach samples, frozen spinach and frozen spinach with added cream were added to the sample set resulting in 7 different sample types. Fresh samples were cleaned and washed with cold tap water before spectral acquisition. Cooking was performed using a steam oven and specific cooking times for each product. Frozen and frozen creamed spinach were defrosted and heated according to the producers' instruction. All samples were brought to room temperature before spectral acquisition. Sample plates were created by adding a random amount of mass from a product-specific range (carrots 20 - 80 g, endive raw 5 - 15 g, endive cooked 15 - 50 g and spinach 20 - 60 g) to the plate. After every five samples, the plate was emptied and cleaned. Product samples were put on the plate off centre and placed under the RGBD camera such that the centre of mass was roughly at a randomly selected rotation. In total, 585 samples plates were prepared, by incrementally adding vegetables to the same plate (i.e. 5 times for each plate). A single RGB and depth image was captured for each sample (Figure 2.1 C-D).

2.3 Determination of reference values of butter and vegetables

The total moisture content of the butter samples was determined by NEN-EN-ISO 3727-1:2001 / IDF 80-1:2001, the total fat content by NEN-EN-ISO 17189 : 2003 / IDF 194 : 2003. For the vegetables, the sample mass was determined using a standard-issue analytical balance.

2.4 Spectral image setup and acquisition

The HSI and RGBD image equipment was mounted in a customized build setup with an approximate distance to the sample of 30 and 40 cm respectively (Figure 2.1 E-F). A standard-issue white glazed ceramic plate (diameter 20 cm) was used for sample presentation. The HSI camera was mounted dead-centre above the sample, whereas the RGBD camera was mounted on a 45-degree angle. For the HSI system, four standard issue halogen lights were installed at each corner around the camera. Lights were directed dead-centre downwards to avoid direct illumination of the reflecting ceramic plate and surfaces of the samples presented. An Effilux bar light (ELB-400SW, Effilux, Les Ulis, France) with a diffuse window was installed underneath the RGBD camera.

An IMEC SWIR Snapscan hyperspectral imaging system (Interuniversity Microelectronics Centre, Leuven, Belgium) with a spectral range of 1117 – 1670 nm equipped with an Optec 16mm F1.7 SWIR lens (Optec S.p.A., Parabiago, Italy). The system was controlled by the IMEC Snapscan software (version 1.3.0.8, IMEC). Before data acquisition, the camera was calibrated using a 95% reflection calibration white standard (WS) tile sized 200 x 200 mm, using an integration time of 2.5 ms. Raw sample spectra were corrected according to **Equation 2.1** to generate corrected diffuse reflection spectra used in subsequent calculations for the 640 x 512 x 108 (x x y x λ) sized hypercubes.

$$Corr_{(x,y,\lambda)} = (Raw_{(x,y,\lambda)} - Dark_{(x,y,\lambda)})/(WS_{(x,y,\lambda)} - DarkWS_{(x,y,\lambda)}) \tag{2.1}$$

RGBD images were acquired using an Intel RealSense D415 depth camera (Intel Corporation, Santa Clara, CA, USA) controlled by a custom build GUI. Images were captured in a semi-controlled environment, with light blocking on the left, right and backside of the scene using mat dark plastic plates. Furthermore, a dark A2-sized paper was used as a background on which plates were placed. The exposure time was set to 70 ms to prevent saturation of the image. Intrinsic calibration parameters available on the camera were stored alongside the color and depth image to be able to construct a point cloud of the depth image.

2.5 Data analysis

Multivariate statistics of HSI data

Processing of the HSI images was performed using the hyperSpec package [2] in R 3.6.1 [3]. The corrected images contained hyper-responsive (dead) pixels and non-sample regions, both of which needed to be excluded from further processing (Figure 2.1 B). Dead pixels contained reflection values $> 1e6$ and were easily recognized and removed. To discriminate between the sample (region of interest, ROI) and the background (empty plate, table and showy edges were present), a simple spectral based principal component analysis (PCA) approach was performed, in which scores on Principal Component (PC) 2 and a threshold of 0 determined if any pixel belonged to the

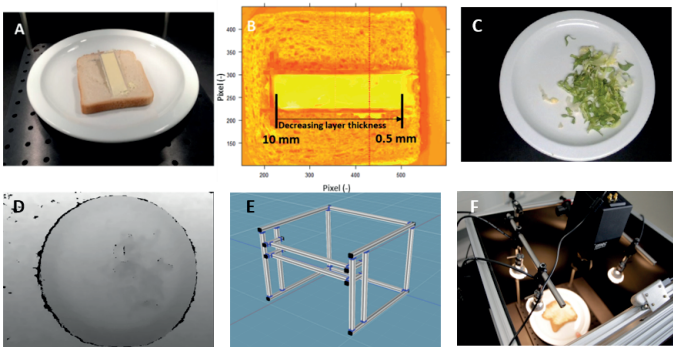

Figure 2.1: (A) Triangular mold for accurate layer thickness estimation of butter; (B) Spectral Image Heat plot of the butter applicator on a wholegrain sandwich; (C) RGB image of raw endive; (D) Depth image of (C) visualized using a gray color map; (E) Frame design of food intake setup; (F) Experimental setup with Hyperspectral SWIR camera (center) and RGBD camera (right bottom).

sample or not. As PCA's signs are non-determined, the portion near the centre of the image – presumed to be sample and not background – determined whether positive or negative PC2 scores contained the ROI. No data preprocessing or further thresholding or area filling was needed nor applied.

To estimate the butter and bread types, and mainly to characterize the butter (fat content, type, fat (un)saturation) and quantity (average layer thickness) of the butter on the bread a straightforward approach was chosen. a more straightforward approach was chosen. First, ROI-spectra were corrected for scattering effects (standard normal variate, SNV). Then the intensity of the signal at 1221 nm was found to correlate well with the amount of butter (fat) [4]. Individual spectra from the ROI pixels were sorted based on the value in this number, and the top and bottom 5% of the spectra were averaged to represent the butter and bread-spectra, respectively. These averaged spectra were input for further (multivariate) classification and regression. Although machine learning approaches per pixel (or pixel cluster) of spectral cube data are a normal approach, this method was not necessary for these relatively simple samples.Bread type (3 types) were classified using the

”bottom 5%” spectra for each of the samples using Soft independent modelling by class analogy (SIMCA), employing cross-validation (25x bootstrapped training sets). Butterfat content (Section 2.3) and butter layer thickness on the bread (5 discrete values, Section 2.2) were fitted using multivariate regression (PLS) using leave-10-out cross-validation. To assess the effect of the interaction of the butter layer thickness on bread on the observed spectra, a custom plexiglass shape was built to help apply butter on bread in known thicknesses. This device was 10 cm long and approximately 2 cm wide, and had a height of 10 mm on one end, linearly decreasing to 0 mm on the other end. When placed on a sandwich, this shape can be filled with butter and smoothed with a knife, after which an HSI will yield spectra with known butter thicknesses on bread -– only marginally affected locally by the foamy structure of the bread, slightly increasing the average thickness of the butter layer.

Analysis of RGBD images

Colour and depth information was used to identify and estimate the mass of vegetables on the plate. Two separate Convolutional Neural Networks (CNNs) were trained on colour images; ResNet50 [5] to classify the vegetable present in the scene and U-Net [6] to semantically segment the scene into pixels belonging to vegetables and the background. The segmentation was used to extract the relevant information from the depth images which were subsequently converted into a point cloud using available intrinsic calibration parameters. From the point cloud the volume was estimated. Finally, linear regression models were constructed for each identified product on the volume to obtain the mass. ResNet was the first network to introduce so called skip connections. These connections allow a layer to learn the identity operator, such that layers can be skipped if they are considered superfluous. U-net is a fully CNN using up sampling layers to obtain the same resolution in the output as layers as the input image. Both networks were trained using PyTorch (version 1.2, www.pytorch.org). The initial learning rate for U-net was upto 0.01, and for ResNet50 to 0.001 as pretrained weights were used for the latter. The learning rate was reduced by factor 10 when no improvement was obtained on the validation set after 200 epoch. Early stopping was applied after 400

epoch of no improvements. Samples were partitioned over a training, validation and test set with each respectively 355, 114 and 116 samples.

3 Results and Discussion

3.1 Bread and butter case using HSI

ROIs were set prior to classification of bread types, fat content estimation and layer thickness determination (Figure 3.1). Bread type classification was performed using SIMCA on the averaged spectra assigned to bread (Table 4.1). While it is clear that the spectra contain sufficient information to reliably distinguish between bread and crispbread underneath the butter layer (regardless of the amount and type of butter applied), the discrimination between white and wholegrain bread was not so clear. Apart from the visual color, their composition (moisture, protein, carbohydrates, fat) is not very different, and the difference in fibers is not strong enough in this part of the near-infrared (NIR) spectrum to make this discrimination reliably in this setup [7].

Table 4.1: Confusion matrix for the classification of bread type.

		Actual		
		White bread	Wholegrain bread	Crispbread
Predicted	White bread	29	17	0
	Wholegrain bread	0	11	0
	Crispbread	0	0	27

The spectra were also used to estimate the fat contents and the (average) thickness of the butter layer applied on the bread. Figure 3.1 gives the predictive capability of the multivariate method applied for both provisions. Note that the fat content of the butter was estimated from all layer thicknesses present in the dataset, and the layer thicknesses were estimated regardless of the butter type and fat content, and these parameters were estimated regardless of the type of bread under the layer. As root mean squared error of predictions (RMSEPs) values of 4.6 (fat w/w %) and 0.056 (mm) were found, which can be considered

quite acceptable for the intended application. Of course, models perform slightly better when separate models are made by determining the bread type and layer thickness first before estimating fat content, or bread type and fat content before estimating layer thickness (data not shown), but these improvements require more data to be calibrated and tested more reliably.

For further development of this model, both in terms of robustness and possibly to reduce the calculation cost (currently approximately 5 seconds per sample on a 2.7 GHz 6-core laptop), it is desirable to understand the interaction that two foodstuffs have on the NIR spectrum as obtained by the hyperspectral camera. Figure 3.1 shows the observed spectra as a function of butter layer thickness as obtained using the butter applicator device. Each line is the average of 20 spectra ($\pm$0.4 mm) around the thickness indicated. Spectra are corrected for scatter-effects (SNV) and transformed using (simplified) Kubelka-Munk (K/S) absorption-over-scattering coefficient-values [8]. It is clear that there are trends in the spectra according to the layer thickness, but there does not seem to be a straightforward solution which explains the trend from 0 to 10 mm of butter on the bread for all wavelengths. The same holds for unprocessed and only scaled data. This implies that the desired information is present in the spectra, but the described multivariate statistics and due calibration are required to extract this information.

3.2 Vegetable case using RGBD imaging

To calculate the volume of vegetables, firstly, PCA was performed on point clouds centered around the origin (Figure 3.2). The vectors found using PCA represent the length, width and height of the three-dimensional objects. Next, points were rotated such that the third principal component was aligned with the vector (0,0,-1). After rotation, the surfaces of the scenes were parallel to the plane z = 0. To compensate for noise in the depth image, z-values of the points were subtracted such that the 98^{th} percentile was at z = 0. Finally, point clouds were subsampled in a uniform voxel grid of fixed dimensions and estimates of volumes were obtained by summing the z-values.

The CNNs were trained on the training set, early stopping was applied using the validation set and the results are reported on the test

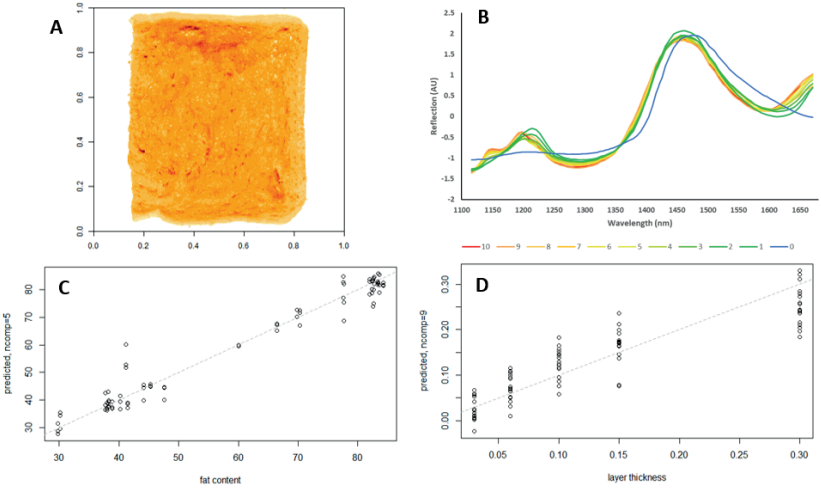

Figure 3.1: (A) A result of ROI detection of the 10 x 10 cm of bread-and-butter detection of HSI data in a false-colour image of first PC scores; (B) Average NIR spectrum as a function of the butter layer thickness (0 – 10 mm); (C) determination of the total fat content of butter independent of the bread type and layer thickness; (D) determination of the layer thickness of the butter independent of the bread and butter type.

set. The classification of the selected vegetables and preparations was successful with zero misclassification as shown in the test set. Indeed, there seems to be sufficient variation present in the RGB data in either color or texture for successful identification. When endive is cooked the product darkens whereas the carrots lose the shiny smooth structure after having been prepared. As for the three different types of spinach, these products themselves are visually different. Estimates of the mass were obtained using product-specific regression models on the calculated volumes. Reasonable results were obtained ranging from an R-squared score of 0.80 to 0.96 (Table 4.2). Although the volume estimation method is in principle generic, as it does not assume any shape or form of the product, to get an accurate mass estimation individual models are needed. A single regression model could be constructed based on the estimated volumes when the (average product) densities are known.

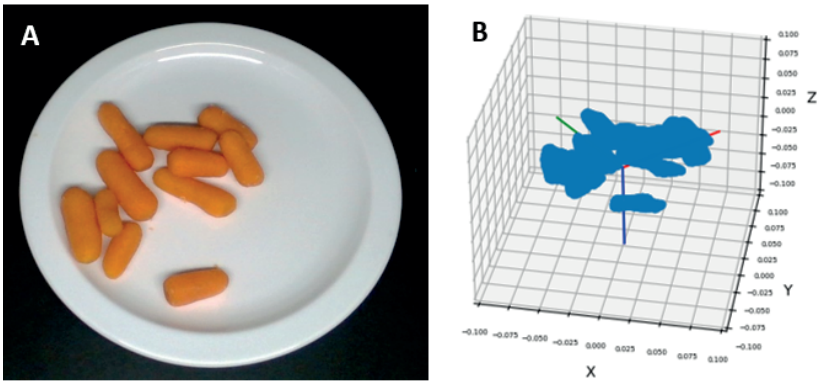

Figure 3.2: (A) Colour image of a scene; (B) Reconstructed point cloud of the segmented vegetables with the first three principal components, corresponding the length (red), width (green), and height or depth (red) axis of the vegetables.

4 Conclusions and outlook

Both sensing systems performed well in the identification and quantification of food products, and the estimation of nutrient composition. To expand the versatility of the setup, a data-fusion experiment will be performed. In this experiment, sandwiches containing multiple elements will be used, such as tomatoes, avocadoes, nuts/seeds, sprouting vegetables and butter. Also, the application of a functional meta-data layer for improving decision making and reducing the number of reference samples will be worked out. In 2021, a human study will be performed to test and validate the setup. Finally, the ultimate goal is to integrate the used analysis techniques into smaller devices such as wearables or smartphones that can be used by consumers themselves. Data obtained by using these wearables can then be combined with data from other sources such as consumer profiles to assess dietary intake as accurately as possible. Missing data can be supplemented based on user profiles, for example, if a particular food is not optimally oriented concerning the measuring equipment or if the food cannot be adequately determined through detection techniques at all. This will be of great value for providing personalized dietary advice that is op-

timally tailored at the individual consumer, eventually contributing to better health.

Table 4.2: R-squared scores for mass estimation of individual vegetables by RGBD imaging.

Food product	R-squared
Spinach creamed	0.96
Spinach frozen	0.93
Spinach fresh	0.88
Endive raw	0.87
Endive cooked	0.80
Carrots raw	0.91
Carrots cooked	0.83

5 Acknowledgements

This research is part of the Dutch Research Council (NWO) program "Dutch National Research Agenda: Start Impulse Measuring and Detecting of Healthy Behavior" under grant number 400.17.604. `www.Metenendetecteren.nl`

References

1. J. Marín, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, and A. Torralba, "Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 187–203, 2021.
2. C. Beleites and V. Sergo, *hyperSpec: a package to handle hyperspectral data sets in R*, 2018, r package version 0.99-20180627. [Online]. Available: http://hyperspec.r-forge.r-project.org
3. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: https://www.R-project.org/

4. F. R. van de Voort, J. Sedman, and T. Russin, "Lipid analysis by vibrational spectroscopy," *European Journal of Lipid Science and Technology*, vol. 103, no. 12, pp. 815–826, 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/1438-9312%28200112%29103%3A12%3C815%3A%3AAID-EJLT1111815%3E3.0.CO%3B2-P

5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv:1512.03385 [cs]*, Dec 2015, arXiv: 1512.03385. [Online]. Available: http://arxiv.org/abs/1512.03385

6. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

7. T. Ringsted, H. W. Siesler, and S. B. Engelsen, "Monitoring the staling of wheat bread using 2d mir-nir correlation spectroscopy," *Journal of Cereal Science*, vol. 75, pp. 92 – 99, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0733521016303812

8. F. Boroumand, J. E. Moser, and H. V. D. Bergh, "Quantitative diffuse reflectance and transmittance infrared spectroscopy of nondiluted powders," *Applied Spectroscopy*, vol. 46, no. 12, pp. 1874–1886, 1992. [Online]. Available: https://doi.org/10.1366/0003702924123502