ORIGINAL ARTICLE

An evaluation of a virtual atlas of portion sizes (VAPS) mobile augmented reality for portion size estimation

Meng Chun Lam[1] · Nur Afyfah Suwadi[1] · Adibah Huda Mohd Zainul Arifien[2] · Bee Koon Poh[2] · Nik Shanita Safii[2] · Jyh Eiin Wong[2]

Abstract

Food portion size estimation is a critical yet challenging task in dietary assessment. Augmented reality technology enables the presentation of food dimensions and volume in a virtual three-dimensional object. It has the potential to improve perception and estimation of portion sizes. This study aims to develop and evaluate a novel mobile augmented reality application, namely Virtual Atlas of Portion Sizes (VAPS), as a portion size estimation aid. The development methodology of VAPS involves food photography, reconstruction of 3D models using photogrammetry method and presenting them in an AR environment. The 3D food models displayed in either semi-transparent or vivid mode for users to perform food portion estimation. Users can then resize and rotate the 3D models to fit the virtual model with the actual food. A total of thirty-six participants were involved in the evaluation and were divided into a health science and a non-health science background group. VAPS received good usability level with 76 SUS score. In terms of task completion time, unsurprisingly, the health science group performed faster. However, both groups have equivalent accuracy on the food portion estimation task using VAPS: 22.5% for non-health science group and 26.6% for health science group. The health science group liked and have better accuracy in vivid 3D food models (37.5%). Meanwhile, the non-health science group preferred semi-transparent 3D food models, but the accuracy is not significantly different between semi-transparent (25%) and vivid 3D food model (20%). Results demonstrate the potential of VAPS to aid in portion size estimation for dietary assessment, and participants' feedback will be incorporated in the future for improvement of the app.

Keywords Food portion · Portion estimation · Food atlas · Augmented reality · Human–computer interaction

1 Introduction

Dietary assessment plays an important role in determining nutritional status of individuals and groups and understanding diet-disease relationships. However, accurate measurement of dietary intake is considered to be an open research problem in the field of nutrition and health (Fang et al. 2016). Typically, dietary intake information is obtained through self-reporting methods, such as 24-h diet recalls, dietary history, diet records, and food frequency questionnaires. While individuals can reliably report the types of food they have consumed, they are less capable of quantifying the amount or portion of foods consumed. As the accuracy of portion size estimates has been reported to be low in previous research (Foster et al. 2006; Nikolić et al. 2018; Ovaskainen et al. 2008), and even among nutrition professionals (Ainaa Fatehah et al. 2018), it is not surprising, therefore, that portion size estimation inaccuracy is deemed as one of the largest measurement errors in most dietary assessment methods (Gibson 2005).

To overcome the challenges associated with portion size estimation, 2-dimensional food photographs are often used in such dietary assessment methods as diet records or 24-h recalls. Portion size photography, either in printed or electronic version, represents the range of small to large portion sizes for commonly consumed food items. They are typically

✉ Jyh Eiin Wong
wjeiin@ukm.edu.my

[1] Mixed Reality and Pervasive Computing Lab, Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[2] Center for Community Health Studies (ReaCH), Faculty of Health Sciences, Universiti Kebangsaan Malaysia, 50300 Kuala Lumpur, Malaysia

presented as a series of photographs depicting graduated sizes for each food item and are bound together in an album format. These food photographs are used as visual prompts and portion size estimation aid to help individuals describe the amounts of food they have eaten. Portion size photography, in either print or electronic version, represents the range of portion sizes of commonly consumed food items. It is normally incorporated into dietary assessment methods, such as dietary records and 24-h recalls, to help individuals describe amounts of food eaten. Individuals can select the food photograph that resembles most closely the amount of food they have actually consumed, and hence more accurately estimate the portion sizes consumed (Martin et al. 2008). Food portion size estimation using food photographs remains a challenging task, as it requires three primary skills—perception, conceptualization and memory (Nelson et al. 1994). Perception refers to an individual's ability to relate an amount of food, which is present in reality, to the amount depicted in a photograph. Conceptualization is the ability of an individual to make a mental construction of a food amount, which is not present in reality, and to relate that construction to a photograph. This process, of course, is affected by the individual's memory (Frobisher and Maxwell 2003). Besides these cognitive skills, estimation of portion size can also be influenced by the characteristics of food. The amounts of some foods, especially foods of indeterminate or amorphous shape, have been reported be more difficult to quantify (Almiron-Roig et al. 2013; Hernández et al. 2006; Hooper et al. 2019).

Augmented reality (AR) is an emerging technology that overlays virtual information on the physical world to enhance the blending between the virtual and physical worlds. AR is able to track the features embedded within an image which acts as a marker (Tan et al. 2018) and renders virtual information, normally a 3D model, on top of it. The 3D model can be interactive, as it enables users to make changes on a model or trigger other functions within an AR app (Muhammad Nizam et al. 2018a, b; Sadik and Lam 2017). Therefore, instead of using printed food photographs in an album or digital photographs in a mobile app, AR is able to show and provide interactable 3D models that enhance the users' ability to visualize the food concerned, including its dimensions and volume.

This study aims to develop a virtual food atlas using AR technology and evaluate its usability, accuracy and gather subjective feedback about its use as a portion size estimation aid. We envision that the invented mobile AR app, namely Virtual Atlas of Portion Sizes (VAPS), can be used to aid portion size estimation, both during (via a food record method) and after a meal (via a 24-h dietary recall interview method). During food recording, users need to perceive the size of the food portion in front of them. Having a semi-transparent overlay allows viewing and matching of the size of the food portion beneath the app. On the other hand, dietary recalls require users to remember and relate the food users had consumed the previous day. Having vivid food models which look similar to real food will enable better conceptualisation of food. Due to different user requirements for different dietary assessment methods, VAPS was designed to include both vivid and semi-transparent food models for estimating food portion sizes. This paper is organized as follows: Sect. 2 discusses related studies of AR apps in nutrition and 3D model generation methods. Sections 3 and 4 outline the development methodology of the VAPS, followed by the experiment setup, and finally Sect. 5 presents the conclusion and directions for future work.

2 Related work

2.1 Augmented reality in nutrition

AR technology is considered a new frontier in the field of human nutrition. Several studies have documented the application of AR technology in nutrition-related work, mainly for changing eating behaviour (Chanlin and Chan 2018), improving nutrition knowledge (Narumi et al. 2012), assisting in portion estimation (Dinic and Stütz 2017; Domhardt et al. 2015; Stütz et al. 2014), and presenting standard portion serving (Rollo et al. 2017).

One of the earliest reported work in this field was aimed at controlling food consumption by changing apparent size of food using AR. Narumi et al. (2012) developed a head-mounted display-based AR, which visualized food portions to be larger than their actual sizes to alter the users' visual perception of the food size. The study demonstrated that the food consumption of participants was affected by the augmented size, leading to less food consumption. Another study applied AR technology in an education context. A mobile AR information system, namely ARFood (Chanlin and Chan 2018), was developed to enhance college students' learning of nutrition knowledge. When triggered by food image scan, ARFood provided nutrition information and enabled live access to nutrient information on a daily basis. Chanlin and Chan (2018) classified the participants into high and low levels of monitoring group. Participants with lower-level monitoring were less engaged or checked for specific information less in the app, while high-level monitoring participants used the app more. The study reported that 40% of the participants were in the low-level monitoring group. Participants in this group felt it was cumbersome to perform daily meal recording and were not keen to use the app extensively. In contrast, participants in the high-level monitoring group (60%) were more engaged with the ARFood app and had improved nutrition knowledge and behaviour. They recorded, checked and monitored their nutrient intake

more often and analysed their dietary intakes to maintain a healthy eating lifestyle.

There are only a few studies that have used AR technology to estimate food portion sizes (Rollo et al. 2017; Stütz et al. 2014). Stütz and colleagues developed Eat$_{AR}$ (Stütz et al. 2014), a mobile AR (MAR) that overlaid a 3D shape, such as hemisphere, on top of a food (such as rice). In Eat$_{AR}$, the 3D hemisphere shape's volume can be adjusted by using either the 3-point user input or a time-based touch gesture to fit over a real food. While the results showed that MAR systems significantly outperformed the eyeballing and hand methods in portion estimation, the participants favoured the time-based touch gesture less. The challenges of using this method were related to food types of different shapes such as ice cones, banana and pasta sauces. It was apparent that Eat$_{AR}$ should support different predefined 3D shapes and incorporate a simpler user interaction to adjust the 3D shapes. Therefore, EatAR Tango was developed in subsequent research with simpler user interaction by utilizing the depth sensor on the smartphone's camera (Dinic and Stütz 2017). EatAR Tango provides an unsorted list of 3D points captured in the presented food scene. A user defines the area of the food portion by removing the non-related 3D points to construct a 3D mesh of the food. Hence, the volume of the food portion can be figured out. Because EatAR Tango relies on users to determine the boundaries of a portion of food, the experiment showed that users might mislabel the area of food resulting in overestimation of food volume (Dinic and Stütz 2017).

Another study by Domhardt et al. (2015) conducted a pilot study of BE$_{AR}$, a mobile phone application using augmented reality technology to help diabetic patients estimate the amount of carbohydrates. The estimation of carbohydrate content was based on the patients' selection of food types and 3D redrawing of said food on the smartphone screen. Patients were instructed to use BE$_{AR}$ in real-life for three weeks and were tested at the beginning and at the end of the study. The researchers found that, at the end of the study, in 44% of the estimates of carbohydrate content, there was an error reduction of 6%. More recently, Rollo et al. (2017) demonstrated the potential of ServAR, a MAR application that displays a virtual standard food serving over a plate. These virtual food images had been modified through photo editing software and acted as standard serving or reference sizes, following the Australian Guide for Healthy Eating. Even though the food was not presented in 3D form, ServAR showed potential as a tool to guide the serving of food for portion control.

Taken together, the application of AR technology in the field of nutrition is fraught with numerous limitations and still requires exploration in different contexts. Previous research has demonstrated the current limitations of the AR technology associated with the supportability of the mobile device (Chanlin and Chan 2018), interaction technique that is yet to be user friendly (Domhardt et al. 2015), requires different approaches to estimate the food portion (Stütz et al. 2014). Specifically related to food portion size estimation, AR-enabled tools including EatAR (Stütz et al. 2014), EatAR Tango (Dinic and Stütz 2017), BE$_{AR}$ (Domhardt et al. 2015) and ServAR (Rollo et al. 2017) are developed for use when real food is in front of users. However, when such a tool is used during dietary recording, it is difficult to view the real food underneath the 2D virtual overlay of vivid food model (Rollo et al. 2017). In this study, we developed VAPS which utilizes AR technology to aid both prospective (diet records) and retrospective (24-h dietary recalls) dietary assessment methods, with or without the presence of food. Our research addresses previous gaps by developing a 3D virtual food atlas with both vivid and semi-transparent food models.

2.2 3D model generating

One of the main elements in VAPS is the 3D model itself. 3D models have been used extensively in industries such as entertainment, education, architecture, medicine and healthcare. 3D models represent data in an intuitive, scalable and easily interpretable way (Cheng 2012). The models are 3D virtual sample of the real-world and offer a clear and intuitive visualization of the elements represented. The creation method of 3D model provides a common and interesting solution for the documentation and visualization of real-life objects. This is due to its remarkable results and its affordability. Different 3D models can be obtained depending on the conditions of the project (time and budget), techniques used, target audience and platform of visualization, among others (Owda et al. 2018). There are a few ways to generate 3D model, which includes using 3D modelling software (Havemann and Fellner 2005), 3D scanning the physical object using 3D scanners (Daneshmand et al. 2018) and photogrammetry method (Chandler and Buckley 2016). In this study, the 3D food model needs to be created before it can be imported to the AR environment. Creating a realistic 3D model using the 3D modelling method requires highly skilled personnel and a lot of time and effort, especially the foods having different shapes, sizes, volumes and textures. 3D scanning has a drawback that interferes with 3D model construction, especially when dealing with physical objects that have dark, shiny or clear surfaces, for example, foods that have been cooked with oil. The shiny surfaces of such foods may disrupt the data acquisition process, causing the 3D scanner to fail in obtaining accurate data to generate the 3D model. After considering all these limitations, the photogrammetry method was used in this study to generate the 3D food models for VAPS.

Photogrammetry uses measurements from photographs to produce drawings, maps, measurements or 3D models of real-world scenes or objects (Photogrammetry 2017). Due to low cost and high accessibility, many 3D models are constructed using this method (Hernandez and Lemaire 2017). Photogrammetry captures and processes multiple images of an object from different positions and angles (Chandler and Buckley 2016). It is a complex process that involves camera calibration and orientation, image point measurements, 3D point cloud generation, mesh generation and texturing to capture high quality images (Remondino et al. 2005). However, even though photogrammetry is a complex process, there are software packages such as 3DF Zephyr (3DFlow 2014) to deal with the 3D reconstruction. Using photogrammetry is inexpensive and does not take much time to produce the required data (Ch'ng et al. 2019; Remondino 2011). Photogrammetry is able to deliver metric and detailed 3D information accurately

and reliably with precise estimation of measured image correspondences (Mikhail et al. 2001).

3 Methodology

Figure 1 illustrates the methodology of the VAPS app development. The three main steps are: (1) the generation of the 3D food model via photogrammetry; (2) refining the 3D model; and (3) creating the AR environment.

3.1 Creation of the 3D food model

In the first step, photogrammetry software 3DF Zephyr (3DFlow 2014) was used to stitch multiple images to construct 3D models. The photographic acquisition phase is critical for construction of the 3D model, as the software, 3DF Zephyr, must have a good dataset of images. Blurred

Fig. 1 Development methodology of VAPS
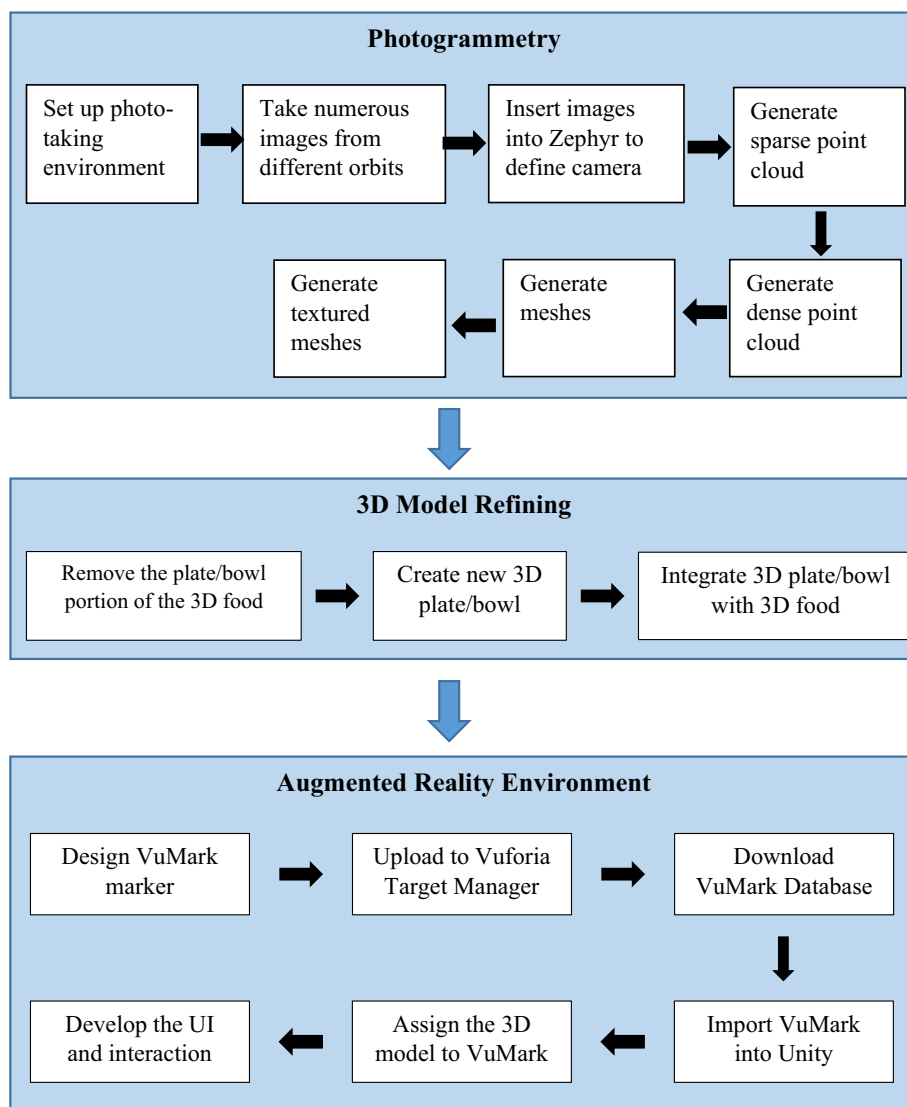
images and dataset with no overlapping pictures are examples of unacceptable data. We used an Asus ZenFone AR with 23-megapixel camera to photograph the food images. A plate or bowl of food was rotated on a 12-inch turntable (Lazy Susan), thus making it easy for photographs to be taken from different angles. To ensure the quality of the lighting, photography for all food items was conducted in a portable photo studio light box. Because of the poor characteristics and repetitive features of the plate and bowl, a piece of flyer paper was put under the bowl/plate to increase the features of the photograph to help Zephyr better recognize the dinnerware. The Asus ZenFone AR was mounted on a tripod in a fixed distance (approximately 25 cm from the edge of the turntable) so that the food would always appear in the centre of the frame. For each food image, a minimum of three orbits were taken with an average of 200 images

each. After the food images were taken, they were then imported into 3DF Zephyr and the generation of sparse point cloud was initiated and follow the official tutorial (3DFlow 2017) to generate the dense point cloud, 3D mesh and texture mesh as shown in Fig. 2.

3.2 Refinement of 3D food model

As shown in Fig. 2b, the 3D food model generated by 3DF Zephyr has some flaws, especially where the plate and bowl (which contained the food) are concerned. These flaws are due to the difficulty photogrammetry has in detecting and capturing shining surfaces and repetitive features. Therefore, the 3D food model must be refined before it is acceptable. Figure 3 shows the refining process. All unnecessary vertexes or polygons were removed using the delete function in

Fig. 2 The results of 3D model generation: **a** sparse point cloud; **b** dense point cloud; **c** 3D mesh and **d** texture mesh
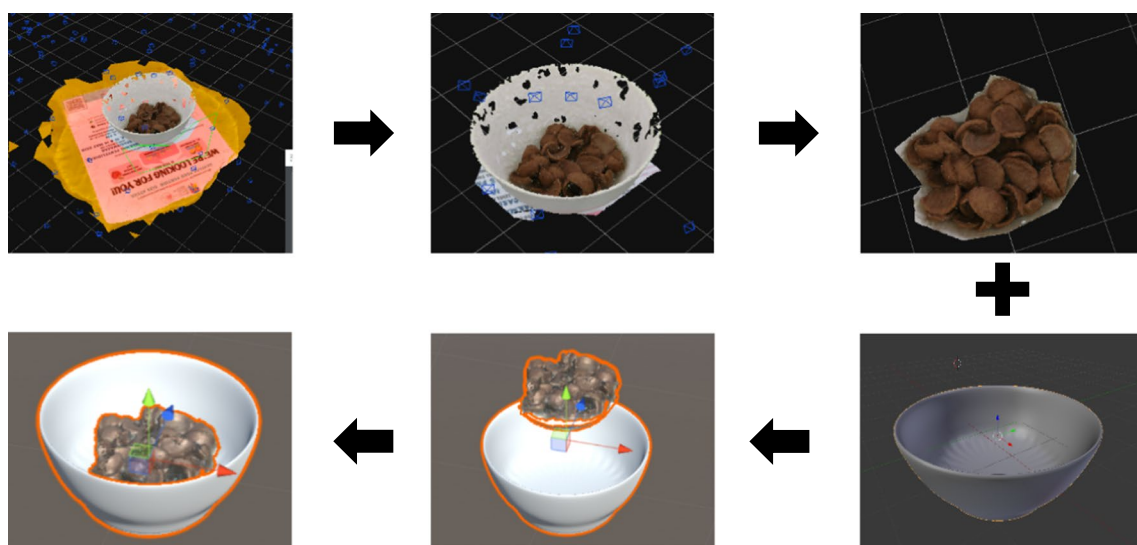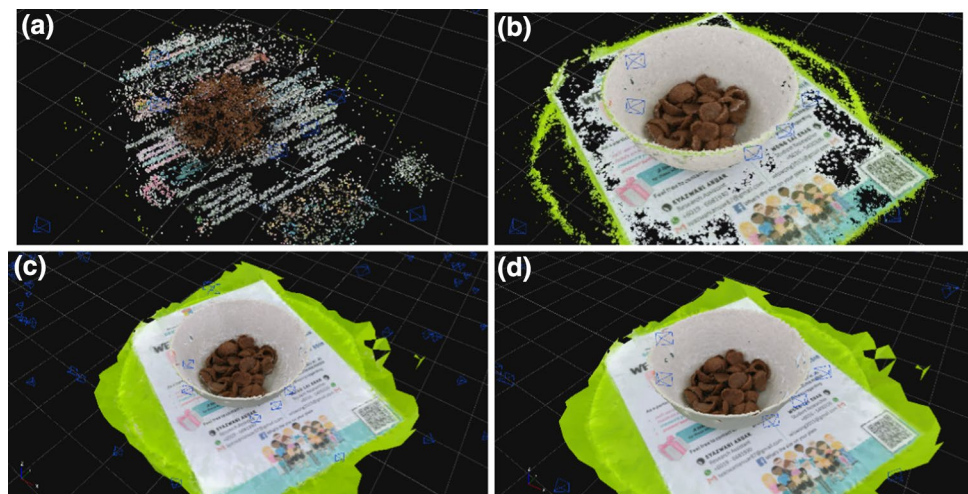

Fig. 3 Result of 3D model refining

3DF Zephyr, so that only food was retained in the 3D model. A 3D bowl and plate were then created, following the actual size of the physical plate and bowl, using Blender 2.79 software (Blender 2017). Eventually, the 3D food and the bowl/plate model were exported to the OBJ format supported by Unity software for the creation of the AR environment. The integration of both models was done in Unity.

3.3 Creation of augmented reality environment

To create an AR application, a tracking target needs to be defined. The AR engine used in this study was Vuforia. In general, Vuforia supports two types of image-based tracking. The first type is any image that is available in real-life, such as movie posters, food packages, or a brand logo. Second is VuMark, a marker that is similar to the QR code which can be customized in different shapes or forms and have a standardized construct for large numbers of AR targets. In addition to being highly customizable to closely reflect a project's brand identity, VuMark also stores encoded data and initializes a smooth tracking experience. VuMark was used because it guarantees a smooth tracking AR experience and support standard design for the AR marker in an AR application. Figure 4 shows the VuMark that has been

Fig. 4 VuMark elements

designed by following the official instruction from Vuforia (2019) and used in this research.

VAPS (Fig. 5) was developed using Unity together with Vuforia. The generated VuMarks were imported into Unity. The VuMarks were then associated with the corresponding 3D food model sets. For example, 3D food models generated for chocolate cereal came in four portion sizes, which weighed 10, 30, 45 and 60 g. The 3D models were rendered in both vivid and semi-transparent forms with 50% transparency. Users can trigger the rendering/display mode via a button on the phone's display. In addition, users can resize or rotate the 3D food model using a rotation button to overlay the 3D food model on the real food displayed in front of them. After that, users can select one food model (one size) best associated with the actual size of the real food.

4 Experimental design

This experiment aimed to examine the usability and accuracy of the newly developed VAPS app, determine the preferences of rendering mode of 3D food models (vivid/semi-transparent) and collect feedback for improving the app. Because VAPS targets amorphous foods (i.e. foods that take the shape of the container they are in), food was presented on typically used dinnerware. Two 3D food models, fried flat noodles and chocolate breakfast cereal, were presented in four portion sizes each. The fried flat noodles was served on a plate, while the chocolate breakfast cereal was served in a bowl.

Participants were invited to attend a 30-min evaluation session in a laboratory setting. The evaluation adopted the "think aloud" method and a System Usability Scale (SUS) questionnaire (Brooke 1996). Each participant was given four tasks (described in the next section) to estimate the portion size of the two foods using the different user interface modes of the AR app. To begin the evaluation, a facilitator welcomed the participants and briefed them about the study and procedure. All participants completed an informed consent form before participating in the study.
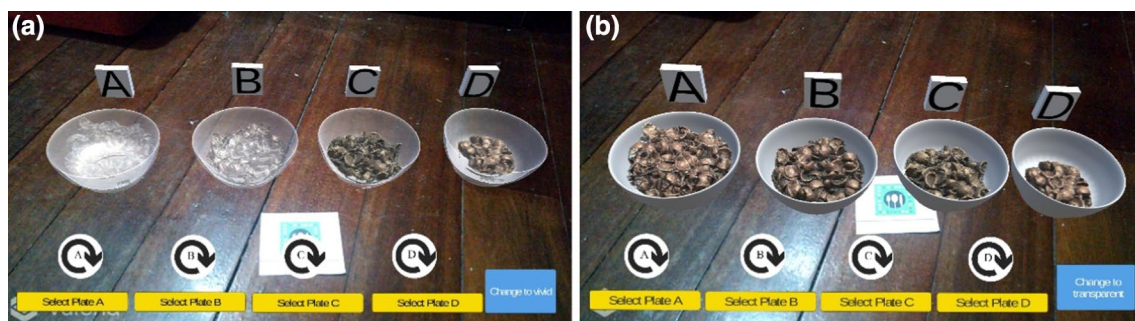

Fig. 5 VAPS—Augmented Reality App, **a** semi-transparent mode and **b** vivid mode

Next, the facilitator explained the "think aloud" method, which encouraged the participants to express their thoughts when performing tasks. The facilitator demonstrated the operation of VAPS to the participants before they started performing the given tasks. The participants did not receive any direct training on the assessment of portion sizes using the VAPS app.

The participants were required to read the task scenario aloud and to begin working on them while thinking aloud. The experiment was set at a table, and the participants were instructed to sit, while performing the tasks, as shown in Fig. 6. The session continued until all four task scenarios were completed, with the order of the tasks shuffled or randomized for each participant. The facilitator took notes of each participant's behaviour, selection (which food model) and completion time for each task. Finally, the participants were asked to complete a 10-item SUS questionnaire on a five-point scale. They were given a small incentive (MYR 10) as compensation for their time. The evaluation session was attended by only one participant at a time, and all materials and equipment were reset at the end of each session.

Figure 7 shows the experimental environment setting for the marker and physical food during the evaluation. Three participants used VAPS to estimate the portion sizes of fried flat noodles and breakfast cereal displayed on the table (Fig. 6). In general, participants were required to use VAPS AR app to select one 3D food model that most closely resembled the real food amount presented. They needed to indicate whether the food amount depicted in the selected 3D food model was *more than*, *less than*, or *the exact same amount* as portion size of the real food presented. The 3D food models in VAPS were presented in both vivid and semi-transparent opacity. The participants could change the 3D food model presentation (vivid/semi-transparent) by tapping a switch button in the app. Both variables were paired, and four (randomized) task scenarios were generated, as described below:
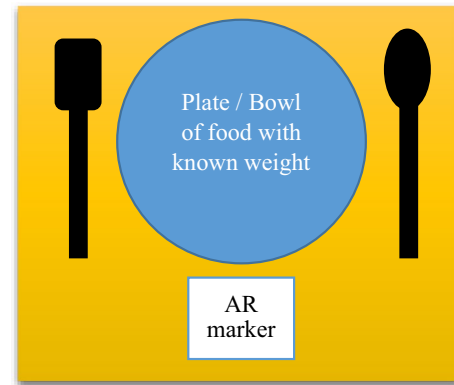

Fig. 7 Experiment table setting for the usability test

Task 1: Select a vivid 3D food model in the VAPS app that is closest to the real food amount presented on the plate (fried flat noodles).

Task 2: Select a semi-transparent 3D food model in the VAPS app that is closest to the real food amount presented on the plate (fried flat noodles).

Task 3: Select a vivid 3D food model in the VAPS app that is closest to the real food amount presented in the bowl (chocolate breakfast cereal).

Task 4: Select a semi-transparent 3D food model in the VAPS app that is closest to the real food amount presented in the bowl (chocolate breakfast cereal).

5 Results and discussion

A total of 36 adults (22 females, 14 males) participated in this experiment. The participants were aged between 21 and 40 years, of which 11% were above 30 years old. More than two-thirds (69%) of the participants were students, while the rest were working adults (31%). The majority of the participants are attending institutions of higher education or

Fig. 6 Participants using the VAPS app during the usability test

have attained tertiary education qualification (81%), with such diverse education backgrounds as information technology, accounting, social sciences, nutrition and dietetics. Participants were categorized as "health science background group" ($n = 16$) if they were studying or working within the field of health sciences including nutrition, dietetics and biomedical sciences. Three types of data were collected during the experiment, which are subjective data (SUS questions, preferences and open questions), accuracy and completion time to estimate the food portion size.

5.1 SUS score

Based on the SUS questionnaire findings and its interpretation (Bangor et al. 2008), VAPS was deemed acceptable and considered a good app by the participants. The SUS score of 76 indicated that VAPS is at the second highest quartile level.

5.2 Accuracy and preference

In terms of accuracy when estimating food portion sizes using the app, three interpretations were recorded—correct, overestimated and underestimated. "Correct" was recorded if the participant selected the exact portion size, "overestimated" if the participant selected a portion size larger than the real portion size, while "underestimated" was recorded when the participant selected a portion size smaller than the real portion size. An adjacent portion size less than 33% error in relation to the real weight is the benchmark used to define "acceptable results" in previous work (Stütz et al. 2014). In this experiment, the mean relative error is 32% between two portion sizes. However, in this study, we do not accept the adjacent portion size selection as a "correct" estimation.

Table 1 shows the overall result of food portion size estimation accuracy for dinnerware and rendering. The results showed that 35 out of 144 (24.3%) portion size selections were correct. Generally, the participants were more likely to overestimate the food portion sizes (46.5%), especially when the food was presented on a plate (76.4%) and in vivid rendering mode (56.9%). When the food was presented in a bowl, 34.7% of the participants selected the correct portion size. Chi-square test was used to determine the significance

of the independent variable because the output data for accuracy consist of three nominal outcomes (Johnson and Karunakaran 2014). An association between accuracy and dinnerware was observed using the Chi-square test ($\chi^2(2) = 52.692$, $p < 0.05$). Chi-square test also reported a significant difference between rendering mode and accuracy, $\chi^2(2) = 13.596$, $p < 0.05$. Vivid 3D model (27.8%) resulted in higher accuracy than semi-transparent 3D model (20.8%). This relatively higher accuracy may be due to the constrained environment (bowl), which allowed participants to better estimate the food amount by comparing the shape of the physical bowl and the top surface of the food with the virtual bowl. On the other hand, the plate is considered as an open environment, where the food is spread out (in width and depth) on the plate's surface; when comparing the physical and virtual food on the plate, the participants had to make a complex comparison of all the sides and top of the food, resulting in less accurate estimation (13.9%).

The results for participants from non-health science and health science backgrounds group are shown in Table 2. Both groups tended to perform better when food was served in a bowl ($p < 0.05$); the non-health science background group achieved 37.5%, while the health science background group had 31.2% accuracy. The results indicate it is more difficult to judge when food was served on a plate, even for the health science background group. Interestingly, for the rendering mode, the results showed that the non-health science background group had equivalent accuracy for the rendering mode, vivid 3D model (20.0%) and semi-transparent 3D model (25.0%). However, the health science background group performed better with the vivid 3D model (37.5%) than the semi-transparent model (15.6%, $p < 0.05$). This result is further supported with their preference for the vivid 3D model, instead of the semi-transparent 3D model, as shown in Fig. 8. In general, accuracy performance for both groups is equivalent ($\chi^2(2) = 1.629$, $p > 0.05$). This suggests that non-health science background participants were able to estimate the portion sizes of food with the same accuracy as health science background participants when using VAPS, although non-health science background did not had any formal food portion size training. Using the strict criteria for correct estimation, our preliminary findings are promising when compared to those reported

Table 1 Accuracy of food portion size estimation by dinnerware and rendering mode

Type	Underestimated n (%)	Correct n (%)	Overestimated n (%)	Significance
Plate	7 (9.7%)	10 (13.9%)	55 (76.4%)	$\chi2(2) = 52.692$, $p < 0.01$
Bowl	35 (48.6%)	25 (34.7%)	12 (16.7%)	
Semi-transparent	31 (43.1%)	15 (20.8%)	26 (36.1%)	$\chi2(2) = 13.596$, $p < 0.01$
Vivid	11 (15.3%)	20 (27.8%)	41 (56.9%)	
Subtotal	42 (29.2%)	35 (24.3%)	67 (46.5%)	

Table 2 Food portion size estimation accuracy for dinnerware and rendering for non-health science and health science background groups

Type	Underestimated n (%)	Correct n (%)	Overestimated n (%)	Significance
Non-health science background group				
Plate	0 (0%)	3 (7.5%)	37 (92.5%)	$\chi2(2)=55.561, p<0.01$
Bowl	21 (52.5%)	15 (37.5%)	4 (10.0%)	
Semi-transparent	11 (27.5%)	10 (25.0%)	19 (47.5%)	$\chi2(2)=0.489, p=0.78$
Vivid	10 (25.0%)	8 (20.0%)	22 (55.0%)	
Subtotal	21 (26.3%)	18 (22.5%)	41 (51.2%)	
Health science background group				
Plate	7 (21.9%)	7 (21.9%)	18 (56.3%)	$\chi2(2)=6.709, p=0.04$
Bowl	14 (43.8%)	10 (31.2%)	8 (25.0%)	
Semi-Transparent	20 (62.5%)	5 (15.6%)	7 (21.9%)	$\chi2(2)=25.611, p<0.01$
Vivid	1 (3.1%)	12 (37.5%)	19 (59.4%)	
Subtotal	21 (32.8%)	17 (26.6%)	26 (40.6%)	
Non-health science background group compared with health science background group				
Non-health science	21 (26.3%)	18 (22.5%)	41 (51.2%)	$\chi2(2)=1.629, p=0.44$
Health science	21 (32.8%)	17 (26.6%)	26 (40.6%)	
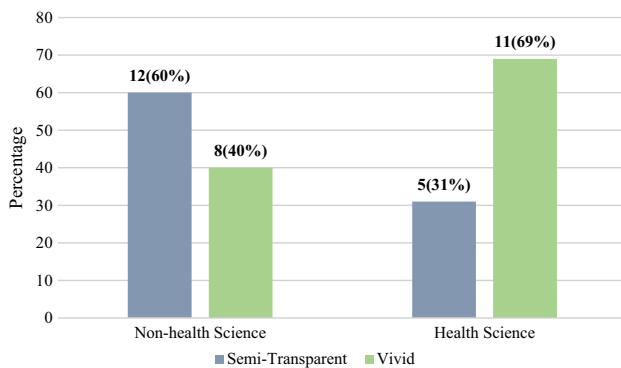
Fig. 8 Preference for rendering mode in VAPS, n(%)

in a recent study (Nikolić et al. 2018) which reported a wide range of correct estimations, ranging from 2.9% (lettuce) to 82.9% (spinach) for 20 selected food items in a sample of nutritional professionals and layperson. Our results demonstrate the potential for an improved version of VAPS to aid portion size estimation, after taking into account users' feedback as discussed in the subjective feedback section.

The preferences of participants were split in half between the two rendering modes, whereby 17 (47%) participants preferred the semi-transparent 3D food model and 19 (53%) participants preferred the vivid 3D food model. There are no outstanding preferences in the rendering mode ($\chi^2(2)=0.111, p>0.05$). However, it is noted that non-health science background participants favoured the semi-transparent models, while health science background participants favoured the vivid 3D food models, as shown in Fig. 8.

5.3 Task completion time

Wilcoxon signed rank test was used to determine the significance of the task completion time within the same background group because the data are continuous, paired sample and not in normal distribution. The data are not normally distributed as the sample size is small, and this is confirmed by the Shapiro–Wilk test with p value less than 0.05. Meanwhile, Mann–Whitney U test was used to test the task completion time between different background group (Johnson and Karunakaran 2014). On average, the participants took 2 min 35 s to complete the portion size estimation task with a standard deviation of 1 min 6 s. Table 3 shows the results of task completion time for the non-health science and health science background groups. The non-health science background group performed faster with bowl ($02:52 \pm 00:57$) and vivid ($02:49 \pm 00:57$) settings. Similarly, the health science background group also performed the portion size estimation faster with the bowl setting ($01:35 \pm 00:22$). However, for the rendering mode, the health science background group performed at similar speed for both semi-transparent and vivid settings, in which Wilcoxon signed rank test showed $Z=-0.944, p>0.05$.

As shown in Fig. 9, the health science background participants performed faster, compared with non-health science background participants. While the health science background group required only 1 min 49 s to complete the task, the non-health science background group took 3 min 11 s. From the tests result, it can be concluded that difference in task completion time between health science and non-health science background group was statistically significant ($U=543.5, p<0.01$). This probably because participants in the health science background group, which included participants with a nutrition or dietetics background, have some

Table 3 Portion size estimation task completion time

Type	Non-health science background: completion time (mean ± SD)	Significance	Health science background: completion time (mean ± SD)	Significance
Plate	03:30 ± 01:07	$Z = -2.593, p = 0.01$	02:04 ± 00:28	$Z = -4.044, p < 0.01$
Bowl	02:52 ± 00:57		01:35 ± 00:22	
Semi-transparent	03:32 ± 01:06	$Z = -3.089, p < 0.01$	01:51 ± 00:27	$Z = -0.944, p = 0.35$
Vivid	02:49 ± 00:57		01:48 ± 00:31	
Overall	03:11 ± 01:05		01:49 ± 00:29	$U = 543.5, p < 0.01$

experience judging portion sizes using such methods as food photograph album. Hence, participants in the health science background group can decide portion sizes faster than those in the non-health science background group.

Participants with health science background preferred vivid food models because they looked highly detailed and real. They found vivid models easier to observe and relate to real food portions as some of the participants had previously used real food and 2D photographs to estimate food portion sizes. They also acted differently from the non-health science background group. Nutritionists from the health science background group, for the most part, were satisfied with just sitting or standing and to immediately observe the food model without attempting to view the food model from another position during the experiment. They were familiar with printed food atlas books, where food models are often shown in 45° angles. Hence, they seemed to stick to angles they had been most often exposed to (45°). It could also be because they were not very well-versed with AR technology. They only physically zoomed in to match the size of the food portion by moving the smartphone towards the food. They rarely resized the 3D model using the app's zoom to ensure the size of the 3D model stayed relatively the same in relation to the size of the real food portion. This helps them

to quickly choose the food model they think most closely resemble the real food.

The non-health science background group took longer to complete the tasks because they were unfamiliar with food portion sizes. Therefore, to estimate the portion sizes, they had to try viewing the models at various angles, e.g. changing position from standing to sitting, zooming in and out of the model several times before they came up with their answers. This behaviour during the experiment contributed to the longer time taken to complete the experiment.

5.4 Subjective feedback

Among participants who preferred semi-transparent mode ($n = 17$), 11 of them thought that the comparison of foods was easier using semi-transparent mode, as exemplified by the quotes, "it is easier for me to see and compare the actual food with the 3D food model in the application" and "it is feasible to use only semi-transparent mode for food in a bowl because we can't actually estimate the amount of food in the bowl when using vivid mode". Some also mentioned "semi-transparent model can overlay the food onto the bowl, so we can see the exact amount when using the device. However, in vivid model, we have to guess, and it will not be more accurate than semi-transparent model".

For participants who preferred vivid rendering mode, their reasons were also related to ease of use for portion comparison and more accurate, real, and clearer displays. Six participants expressed that it was easier to compare: "closest appearance to the real food so we can estimate easily"; "vivid model is easier to be compared with the real food because it has solid colour. However, semi-transparent model is also helpful and complements the vivid model because it can partly show the real food under the model, so we can confirm the portion of the food"; "because it is more accurate in visualizations when choosing the correct food portion size"; "easier to do assumptions". Participants also commended the 3D food models in the vivid mode which appeared real: "can see the real dimension of the food item in detail, thus easier to estimate food portion"; "It looks cool and clear. Its appearance mimics the real food".
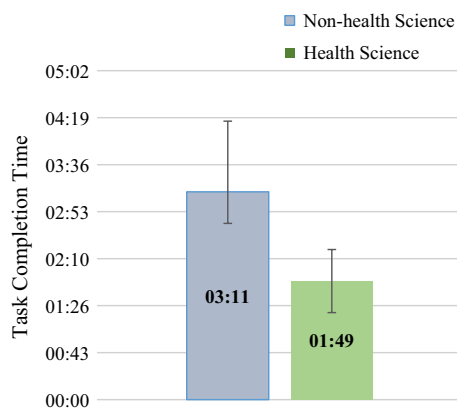

Fig. 9 Task completion time for non-health science and health science background groups

The participants provided suggestions to improve the app at the end of the evaluation. Some of their suggestions were directed at improving the user interface design, adjusting the semi-transparent model to become clearer, including calorie calculation, making the app more interactive and reducing lag. Participants also suggested the addition of a slider to adjust the model's transparency— "the app should allow the user to adjust the vividness or transparency of the AR model. This can improve the accuracy of the portion estimations by letting users adjust the model based of their preferred perspective". The majority of the participants suggested a more attractive user interface to motivate app use, such as "use green colour instead of red to show that the answer is correct", and "user interface can be more interactive (animation) and make it more colourful". Some participants also looked forward to the app having more functions, such as calorie or food portion intake record function, which would allow them to monitor their daily dietary intakes. The app suffered from lag issues which stemmed from our use of highly detailed 3D models to visualize food. Therefore, it was recommended that VAPS should only be run on high-performance smartphones to enable stable tracking and visualization for a smooth AR experience. This lag issue can also be rectified by optimizing the 3D models for use in an AR environment.

Considering subjective feedback from the participants, the key areas to make the app easier to use are: (1) The user interface (UI) can be improved to facilitate smooth user navigation and interactivity (Nikou and Economides 2017). As aesthetics affected user experience positively (Lynch 2009), VAPS could be designed with a clean and consistent colour scheme to increase its attractiveness and consequently encourage users to use the app. Correct answer choices should be indicated using green and positive ping sound effects, whereas wrong answer choices should be indicated using red colour and negative buzzer sounds. (2) A separate training module can be developed to train the user to use the app effectively to estimate food portion sizes. A user should complete the training module first before making a portion size estimation using the app. (3) In the training module, reinforcement can be used to enhance learning. (4) In the current app, the food models are presented in a bowl or plate only, without cutlery as a size reference. In future, commonly used cutlery, such as spoon and fork, should be augmented beside each food model. The addition of cutlery may enhance a user's ability to estimate portion size more accurately by providing a relatable scale of food model in the AR environment with food present in reality (Nelson and Haraldsdóttir 1998).

6 Conclusion

This study applied AR technology in the field of nutrition to help with the estimation of food portion sizes. The main study outcome was the development of the VAPS app which presented virtual 3D food models and then blended them onto the real-world environment. The VAPS app harnesses AR technology to enable users to compare the real food portion size with the virtual 3D food portion size. In the evaluation, VAPS offered two rendering modes (vivid and semi-transparent) to determine which mode was preferred. The results showed both rendering modes receiving positive comments from the participants as each mode had its advantages. Hence, further development of the app will retain both rendering modes as they complemented each other in portion size estimation. All user feedback gathered will be used to improve the design of the app. The improved version of VAPS will be evaluated for its validity in improving portion size estimation relative to other tools to estimate portion size aid during dietary assessment. We hypothesize that the VAPS app enhances visualization and perception of portion sizes, thereby reducing errors in portion size estimation, which is an area of dietary assessment requiring much improvement. VAPS should also be promoted as a learning tool to improve portion size estimation in order to encourage self-monitoring of portion control and healthy eating behaviour.

The VAPS app aims to aid in both prospective (dietary records) and retrospective (24-h dietary recalls) dietary assessment methods. At this stage, we have only tested two types of food as proof-of-concept to explore issues with using VAPS to improve it further. VAPS assumed superior to 2D photographs as the 3D food format provides additional size information. After future validation studies comparing the VAPS app with conventional methods, the generalizability of our findings can be ascertained.

References

3DFlow (2017) 3Dflow Academy—Videotutorial 1-Basic 3DF Zephyr workflow - photogrammetry basics. https://youtu.be/17UTelgZqBg. January 2019

3DFlow (2014) 3DF Zephyr. https://www.3dflow.net/3df-zephyr-photogrammetry-software/. Accessed 28 Oct 2020.

Ainaa Fatehah A, Poh B, Nik Shanita S, Wong J (2018) Feasibility of reviewing digital food images for dietary assessment among nutrition professionals. Nutrients 10:984. https://doi.org/10.3390/nu10080984

Almiron-Roig E, Solis-Trapala I, Dodd J, Jebb SA (2013) Estimating food portions Influence of unit number, meal type and energy density. Appetite 71:95–103. https://doi.org/10.1016/j.appet.2013.07.012

Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. Int J Human-Computer Interact 24:574–594. https://doi.org/10.1080/10447310802205776

Blender (2017) Blender 2.79. https://www.blender.org/download/releases/2-79/. Accessed 15 June 2018

Brooke J (1996) SUS-A quick and dirty usability scale. Usability Evaluation Industry 189:4–7

Ch'ng E, Cai S, Zhang TE, Leow F-T (2019) Crowdsourcing 3D cultural heritage: best practice for mass photogrammetry. J Cultural Heritage Manag Sustain Dev 9:24–42. https://doi.org/10.1108/JCHMSD-03-2018-0018

Chandler JH, Buckley S (2016) Structure from motion (SFM) photogrammetry vs terrestrial laser scanning. IN: Carpenter, M.B. and Keane, CM (eds.) Geoscience Handbook 2016: AGI Data Sheets, 5th ed. Alexandria, VA: American Geosciences Institute, Section 20.1. https://hdl.handle.net/2134/21673

Chanlin L-J, Chan K-C (2018) Augmented reality applied in dietary monitoring. Libri 68:137–147. https://doi.org/10.1515/libri-2017-0024

Cheng H-M (2012) The workflows of 3D digitizing heritage monuments. In: Apolinar Munoz J (ed) Laser scanner technology. IntechOpen, London

Daneshmand M et al. (2018) 3D scanning: a comprehensive survey arXiv preprint arXiv:180108863

Dinic R, Stütz T (2017) EatAR tango: results on the accuracy of portion estimation. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct), IEEE, pp 284–287

Domhardt M et al (2015) Training of carbohydrate estimation for people with diabetes using mobile augmented reality. J Diabetes Sci Technol 9:516–524. https://doi.org/10.1177/1932296815578880

Fang S, Zhu F, Jiang C, Zhang S, Boushey CJ, Delp EJ (2016) A comparison of food portion size estimation using geometric models and depth images. Proceedings international conference on image processing 2016:26-30 doi:10.1109/ICIP.2016.7532312

Foster E, Matthews JN, Nelson M, Harris JM, Mathers JC, Adamson AJ (2006) Accuracy of estimates of food portion size using food photographs–the importance of using age-appropriate tools. Public Health Nutr 9:509–514. https://doi.org/10.1079/PHN2005872

Frobisher C, Maxwell S (2003) The estimation of food portion sizes: a comparison between using descriptions of portion sizes and a photographic food atlas by children and adults. J Hum Nutr Diet 16:181–188. https://doi.org/10.1046/j.1365-277X.2003.00434.x

Gibson RS (2005) Principles of nutritional assessment. Oxford University Press, USA

Havemann S, Fellner DW (2005) Generative mesh modeling. Libri 68:137–147

Hernandez A, Lemaire E (2017) A smartphone photogrammetry method for digitizing prosthetic socket interiors. Prosthet Orthot Int 41:210–214. https://doi.org/10.1016/j.jfca.2006.02.010

Hernández T et al (2006) Portion size estimation and expectation of accuracy. J Food Compos Anal 19:S14–S21. https://doi.org/10.1016/j.jfca.2006.02.010

Hooper A, McMahon A, Probst Y (2019) The role of various forms of training on improved accuracy of food-portion estimation skills: a systematic review of the literature. Adv Nutr 10:43–50. https://doi.org/10.1093/advances/nmy060

Johnson LR, Karunakaran UD (2014) How to choose the appropriate statistical test using the free program "statistics open for all"(SOFA). Annal Commun Health 2:54–62

Lynch P (2009) Visual decision making. A List Apart: For People Who Make Websites, 286. http://www.alistapart.com/articles/visual, Accessed from 02 Feb 2017

Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD (2008) A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 101:446–456. https://doi.org/10.1017/S0007114508027438

Mikhail EM, Bethel JS, McGlone JC (2001) Introduction to modern photogrammetry. Wiley, New York

Muhammad Nizam SS, Abidin RZ, Hashim NC, Lam MC, Arshad H, Abd Majid NA (2018) A review of multimodal interaction technique in augmented reality environment. Int J Adv Sci Eng Inf Technol 8:1460–1469

Muhammad Nizam SS, Lam MC, Arshad H, Suwadi NA (2018) A scoping review on tangible and spatial awareness interaction technique in mobile augmented reality-authoring tool in kitchen. Adv Multimed. https://doi.org/10.1155/2018/5320984

Narumi T, Ban Y, Kajinami T, Tanikawa T, Hirose M Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality. In: Proceedings of the SIGCHI conference on human factors in computing systems, 2012. ACM, pp 109–118. doi:https://doi.org/10.1145/2207676.2207693

Nelson M, Atkinson M, Darbyshire S (1994) Food photography I: the perception of food portion size from photographs. Br J Nutr 72:649–663. https://doi.org/10.1079/bjn19940069

Nelson M, Haraldsdóttir J (1998) Food photographs: practical guidelines II. Development and use of photographic atlases for assessing food portion size. Public Health Nutr 1:231–237

Nikolić M, Milešević J, Zeković M, Gurinović M, Glibetić M (2018) The development and validation of food atlas for portion size estimation in the Balkan region. Front Nutr. https://doi.org/10.3389/fnut.2018.00078

Nikou SA, Economides AA (2017) Mobile-based assessment: Investigating the factors that influence behavioral intention to use. Comput Educ 109:56–73

Ovaskainen M et al (2008) Accuracy in the estimation of food servings against the portions in food photographs. Eur J Clin Nutr 62:674. https://doi.org/10.1038/sj.ejcn.1602758

Owda A, Balsa-Barreiro J, Fritsch D (2018) Methodology for digital preservation of the cultural and patrimonial heritage: Generation of a 3D model of the church St. Peter and Paul (Calw, Germany) by using laser scanning and digital photogrammetry. Sens Rev 38:282–288. https://doi.org/10.1108/SR-06-2017-0106

Photogrammetry (2017) What is photogrammetry? https://www.photogrammetry.com/. Accessed 25 April 2018

Remondino F (2011) Heritage recording and 3D modeling with photogrammetry and 3D scanning. Remote Sens 3:1104–1138. https://doi.org/10.3390/rs3061104

Remondino F, Guarnieri A, Vettore A 3D modeling of close-range objects: photogrammetry or laser scanning? In: Proc. SPIE 5665, Videometrics VIII, San Jose, California, United States, 17 January

2005. International Society for Optics and Photonics, p 56650M. https://doi.org/10.1117/12.586294

Rollo ME, Bucher T, Smith SP, Collins CE (2017) ServAR: An augmented reality tool to guide the serving of food. Int J Behav Nutr Phys Act. https://doi.org/10.1186/s12966-017-0516-9

Sadik MJ, Lam MC (2017) Stereoscopic vision mobile augmented reality system architecture in assembly tasks. J Eng Appl Sci 12:2098–2105. https://doi.org/10.36478/jeasci.2017.2098.2105

Stütz T, Dinic R, Domhardt M, Ginzinger S Can mobile augmented reality systems assist in portion estimation? A user study. In: 2014 IEEE international symposium on mixed and augmented reality-media, art, social science, humanities and design (ISMAR-MASH'D), 2014. IEEE, pp 51–57. doi:https://doi.org/10.1109/ISMAR-AMH.2014.6935438

Tan SY, Arshad H, Abdullah A (2018) An efficient and robust mobile augmented reality application. Int J Adv Sci Eng Inf Technol 8:1672–1678. https://doi.org/10.18517/ijaseit.8.4-2.6810

Vuforia (2019) Designing a VuMark in Adobe Illustrator. https://library.vuforia.com/content/vuforia-library/en/articles/Solution/Designing-a-VuMark-in-Adobe-Illustrator.html. July 2019