IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 10, 2022, accepted April 22, 2022, date of publication May 2, 2022, date of current version May 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3171584

Food Volume Estimation by Integrating 3D Image Projection and Manual Wire Mesh Transformations

SHAMUS P. SMITH[1], (Senior Member, IEEE), MARC T. P. ADAM[1], GRACE MANNING[2], TRACY BURROWS[2], CLARE COLLINS[2], AND MEGAN E. ROLLO[2]

[1]School of Information and Physical Sciences, The University of Newcastle, Australia, Callaghan, NSW 2308, Australia
[2]Priority Research Centre for Physical Activity and Nutrition, School of Health Sciences, The University of Newcastle, Australia, Callaghan, NSW 2308, Australia

Corresponding author: Shamus P. Smith (shamus.smith@newcastle.edu.au)

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation OPP1171389. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Newcastle Human Research Ethics Committee under Application No. HREC #H-2018-0271, and performed in line with the National Statement on Ethical Conduct in Human Research, 2007.

ABSTRACT 2D images can be used to capture food intake data in nutrition studies. Estimates of food volume from these images are required for nutrient analysis. Although 3D image capture is possible, it is not commonplace. Additionally, nutrition studies often require multiple food images taken by non-expert users, typically collected using mobile phones, due to their convenience. Current 2D image to 3D volume approaches are restricted by the need for prescribed camera placement, image metadata analysis and/or significant computational resources. A new method is presented combining 2D image capture and automated 3D scene projection with manual placement and resizing of wire mesh objects. 2D images, with a reference object, are taken on low specification mobile phones. 3D scene projection is calculated by twinning a cuboid in 3D space to the reference object in the 2D image. A manually selected 3D wire mesh object is then positioned over the target food item and manually transformed to improve accuracy. The virtual wire mesh object is then projected into the 3D scene and the volume of the target food item calculated. The whole process is computationally light and runs in real-time as an app on a standard Apple iPad. Based on a user study with 60 participants, experimental evaluations of volume estimates over regular shape and ground truth food items demonstrate that this approach provides acceptable accuracy. We demonstrate that the accuracy of estimates can be increased by combining multiple independent estimates.

INDEX TERMS Food volume estimation, 3D image projection, food analysis.

I. INTRODUCTION

Photographs, typically as 2D digital images, are a common way to collect information on dietary intake [6], [9]. For studies in natural settings, participants often take photographs before, during or after meals. These photographs represent food consumption and are then analysed to classify eating behaviour by, for example, food types and nutritional intake [1].

However, most nutrient analysis of food requires the weight of food items, in addition to food item identification, in order to determine relevant nutritional metrics [14]. Weight is determined by food density and volume. Food density is often determined by food identification and application of standard food density tables which are only available for some foods [15]. However, volume calculations from photographs, as 2D images, is non-trivial without 3D depth information [6]. 3D imaging from common image sources, i.e. mobile phones with multiple front facing cameras to obtain stereo images, is not yet common place or affordable [17]. Thus, there is ongoing research into developing methods to generate accurate 3D volume estimates from 2D images [10].

To support food volume estimation, two main approaches have been applied; with [14], [17] and without [18] reference objects in images. Reference objects can be explicitly placed before image capture, for example a card [17], piece of local

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma.

currency or a standard sized object like a Rubik's cube [14], or extracted from identified image elements, for example plates or serving vessels of known dimensions [2]. Reference objects provide 3D cues to determine image depth, which is needed to generate 3D volume estimates [6]. Alternatively approaches without reference objects attempt to (i) identify image features, for example parallel elements, that can be used to determine an image vanishing point or (ii) use properties of the camera placement and environment knowledge [18], [22], i.e. the camera becomes its own reference object.

Two challenges with approaches using reference objects are that they can be computationally expensive, for example involving use of neural networks [14], [17] or require multiple food images [6]. Another significant issue, for approaches with and without reference objects, is the casual nature of the 2D image capture. In practical use, cameras are not typically on tripods where orientation is known or positions accurately captured. Recent work [22] proposed a hybrid approach with mobile phones grounded on an image plane, i.e. a table, and camera orientation data presented as machine readable data in the image capture. However, this approach restricts camera placement and in many environments where food intake data would be collected, i.e. in rural locations in the developing world, there may be no table or stable surface to ground the camera during eating occasions. Also strict camera placement requirements might be inconvenient for the user or the user might just forget to have strict placement for every image capture. Thus there is no way of being 100% sure that an image was taken on an appropriate surface.

In the approach presented here, a new method is presented where only 2D images with no camera or image meta-data is needed for the 3D food volume estimates. Through a combination of automated and manual activities, the volume of food items in 2D images can be generated in a process that is computationally light and runs in real-time. This offers considerable flexibility in its application with real world image collection where precise camera view information cannot be collected, for example when photographs are taken with low specification cameras/mobile phones or by untrained users under various lighting and photograph framing conditions.

This paper has three main contributions:

1) A new semi-automated method combining 2D image capture and automated 3D scene projection with manual placement and resizing of wire mesh objects to generate volume estimates of food items.
2) An evaluation of the method on representative food item images, where individual volume estimates are collected via a user study (n=60).
3) A demonstration of how combining pair and triple food volume estimates can improve the accuracy of the volume estimates depending on the regularity of the food item shape.

II. BACKGROUND

Portion size estimation has a large and growing body of research. Here, we overview a selection of particularly relevant related works. Recent reviews in this area can be found in [1], [5], [11], [12], [15], [21].

Approaches for single view image-based food portion estimation typically involve the identification of reference objects in the scene. Common approaches include (i) the inclusion of the reference object [6], [17], for example a fiducial, at image generation [7], (ii) use of prior knowledge of objects in the scene, for example known bowl or plate sizes [2], [7], or (iii) knowledge of the capture device, e.g. the digital camera, and environment context [18], [22].

Fang *et al.* [7] use a fiducial marker to estimate the scale and pose of objects in a scene and also provide estimates of camera parameters. Their approach is based on the use of pre-determined geometric models, namely cylinder and prism models, to determine height and radius estimates. Prior knowledge of the container shape is used as geometric contextual information. For example, the prism method utilises the area of the plate in the image. They used their previously determined criteria of 15% error or less [13] and found that out of 19 food types, only 3 (lettuce, French dressing and ketchup) where outside the 15% error range.

Beltran *et al.* [2] consider the use of configurable wire mesh overlays on 2D images to estimate food portion sizes. Eleven wire frames are provided including cuboid, cylinder, sphere, wedge, ellipse, half spheres (bowl and dome), half ellipse, section of sphere (cap), tunnel and irregular shape. After placement, the user can customize each wire frame to food portions using virtual pressure points to change the wire frame's shape. In order to generate reference points, the diameter and depth of standard plates, bowls and glasses, as present in the 2D images, were measured and provided to the portion calculating algorithm.

Beltran *et al.* [2] also considered rater reliability of portion estimates across 150 food images with results from two dietitians and three engineers. They found that although dietitians had high inter-rater reliability for volumes served (r=0.771), this decreased with smaller portions, i.e. for portions left after serving/eating (r=0.629). This was also the case for the engineers but overall the engineers had better scores, although they note this is likely due to familiarisation with the approach. Individual food types were not defined but serving containers were limited, noted as small bowl (n=48), large bowl (n=42) and plate (n=56).

Puri *et al.* [18] present a system that given a set of three images and a verbal description of food items performs object recognition and 3D reconstruction to estimate food volumes. 3D models of the food items are constructed via a pairwise classification framework. Camera pose estimation treats the camera as the reference object. However, the image analysis requires the use of a remote processing site with server-based computer vision processing. Thus, not practical for real-time use or where there is limited internet connectivity.

TABLE 1. Summary of food volume estimation methods.

Reference	Images needed	Uses reference object	Image capture constraints	Processing features
Dehais et al. [6]	Two	Yes. Credit card sized card.	Top view image. Elliptical shaped plate with flat bottom.	Prior dish and food item segmentation.
Fang et al. [7]	One	Yes. Checker-board card.	Prior knowledge of container shape as geometric contextual information.	Prior food segmentation.
Liu et al. [14]	One	Yes. Rubik's cube.	Side view.	Pre-trained neural networks for object detection and food volume prediction.
Okamoto & Yanai [17]	One	Yes. Pre-registered object (card or wallet).	Background of image cannot be textured.	Pre-trained neural network used for food classification.
Puri et al. [18]	Three	Yes. Checker-board card.	Speech data listing food items in image.	Cloud-based server for image processing.
Yang et al. [22]	One	No physical reference object. Use of a virtual reference object of unit size.	Smartphone needs to be set on a level surface tabletop during picture capture.	Use of the smartphone motion sensor for capturing camera orientation.
Our method	One	Yes. Checker-board card.	Top view image. No constraint on plate size or type. Only needs captured 2D image.	No need for prior food/dish segmentation. No prior processing needed. No constraint on image background.

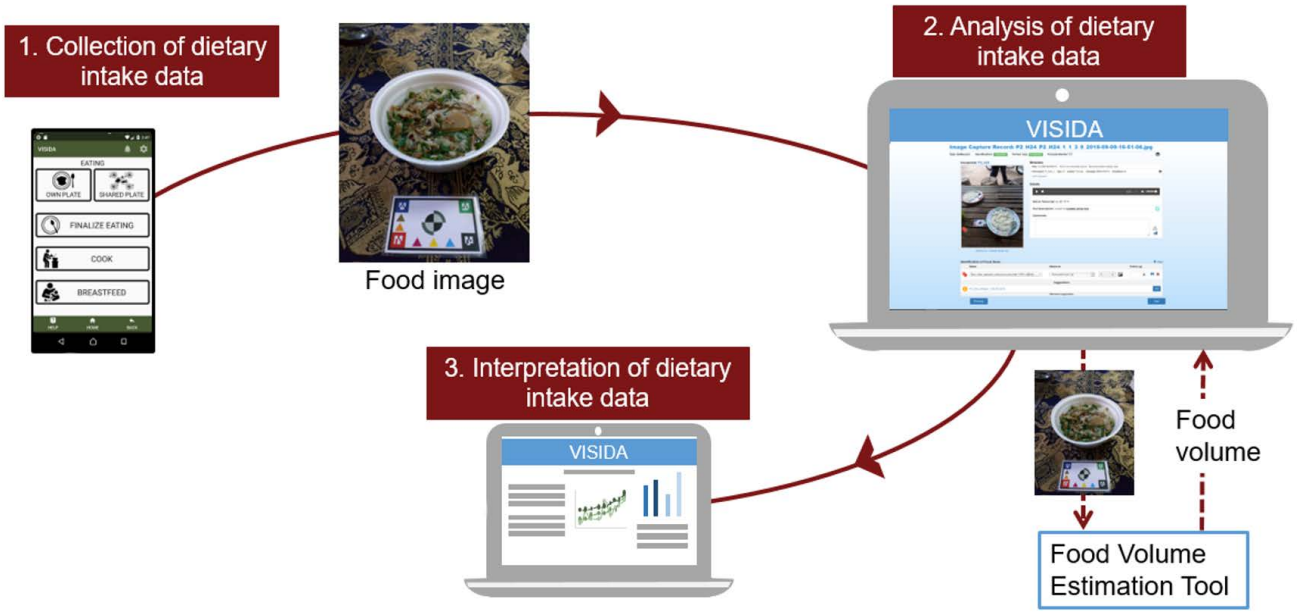

FIGURE 1. Simplified overview of VISIDA system.

Table 1 provides a summary of approaches to food volume estimation, comparing key aspects of each approach. Our proposed method is included for comparison.

III. PROPOSED METHOD

The work described here is part of a larger system which aims to provide a platform for the collection of dietary intake data, including images of food, the analysis of dietary intake data and the interpretation of dietary intake data. A simplified view of the Voice-Image-Sensor technologies for Individual Dietary Assessment (VISIDA) system is shown in Fig. 1. The basic process is that users collect images and voice recordings of food intake data for themselves and others in their household (as required), before and after food consumption, via a smartphone app. This data is uploaded into a content management system (CMS), seen as Step 2 of Fig. 1, where individual food items are identified and matched to food composition databases (i.e. the food items are "tagged"). These tagged images are then queued for food volume estimation and supporting this estimation is the approach reported in this paper. The estimated food item volume is returned to the CMS and contributes to the ongoing analysis of dietary intake data. Here, we present a novel approach to the 3D volume estimation component of this process.

An overview of our approach can be seen in Fig. 2. In Fig. 2, the corners of a reference card on a captured 2D photograph are automatically identified (the blue dots in image 1). A 3D cuboid (the blue rectangle in image 2) is twinned to the reference object and the projection depth angle into the 3D scene determined (the blue arrow in image 2).

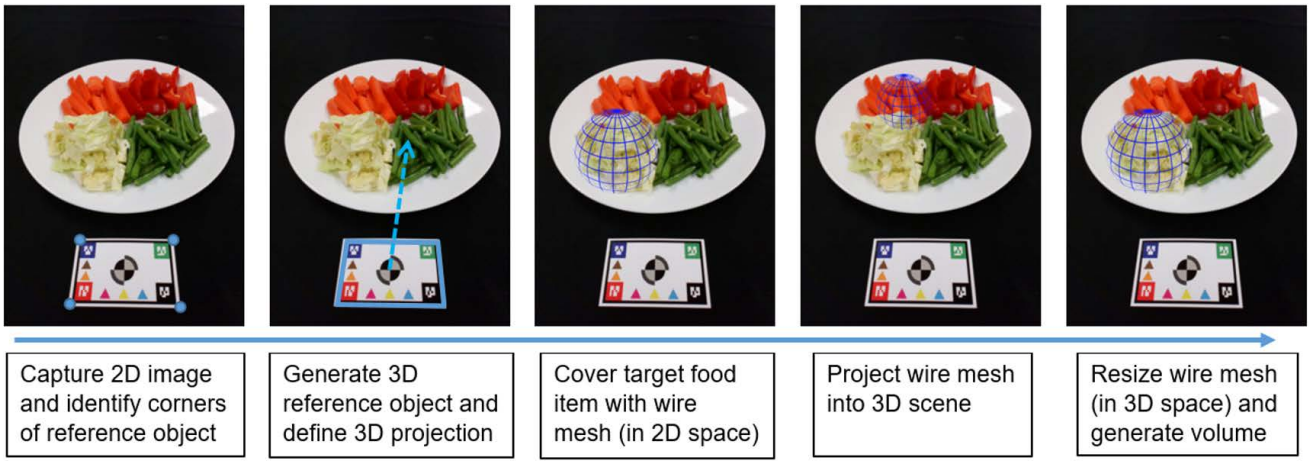

FIGURE 2. Overview of the food item volume estimation approach.

A user selected wire mesh (in this case, a sphere in image 3) is then placed on a target food item. When the wire mesh object is projected into the 3D scene (image 4) it needs to be rescaled to cover the target food item (image 5). With the 3D wire mesh at the correct 3D depth and scale, the volume of the food item it covers can be generated based on the 3D wire mesh shape.

Two examples of the volume estimation tool in use can be seen in Fig. 3 and a breakdown of the manual and automated elements of the approach is presented in Fig. 4. A food image is provided to the tool, running on an Apple iPad, from the CMS where the food item of interest has been tagged. The CMS also provides initial estimations of the fiducial corner tags with the food image. The identification of the fiducial card's corners are automatically determined using the EMGU (v3.4.1) C# adaptation for OpenCV (v4.2.0). As image quality and lighting conditions can impact the automatic corner detections, a manual review of the corner matching is needed (Step 2a in Fig. 4). If the corners are a poor match (as indicated by the tool as an error distance from the known size of the fiducial object), the user can adjust the corner placement manually (Step 2b). As the dimensions of the real fiducial object are known, an equivalent 3D virtual version of the fiducial object is placed at the center of real fiducial object, on the 2D image, and automatically rotated around X, Y, and Z axis to optimise the corners of the virtual fiducial corners to the real fiducial corners (see Algorithm 1[1]). The corner optimisation attempts to find a best-fit match in one dimension at a time, i.e. starting by rotating in the X axis. The rotation algorithm looks ahead in unit rotation increments to determine the best solution at the current angle and will roll back its current rotation if the solution is not improved. This is repeated across the other dimensions, Y and Z. This whole process is then repeated, with the X, Y and Z axis, until no better solutions are found and the mapping has stabilised. Thus, the optimised position of the virtual fiducial is a mapping to the real fiducial card position and enables 3D image depth projection in latter steps.

FIGURE 3. Volume estimation tool example using a dome wire mesh (left) and a box wire mesh (right). A rectangular reference object, a fiducial card, is also included in each captured image.

The user then picks a wire mesh object (sphere, dome, bowl or box) that best matches the food item (step 4) and then resizes and reshapes (X, Y, Z transformations) manually to best match the shape of the food object (step 5). Using the orientation of the virtual fiducial object, the wire mesh object is projected, as a 3D object, into the food image scene (step 6 and see Algorithm 2). This often results in a change of scale of the wire mesh object as it moved into the 3D scene. Thus the user must resize it at this new 3D depth, so that it matches the food item (repeating steps 5-7). When this is complete the wire mesh object is at the projected depth of the food item and its volume can be used to estimate the food item it has covered (step 8 and see Algorithm 3).

To determine the accuracy of the approach, a user study was conducted. The aim of this study was to evaluate the

[1]This is simplified pseudo code for the 3D cuboid orientation optimisation. The full algorithm also includes a decrement step across the Euler angles to avoid local maxima effects on each individual orientation axis.

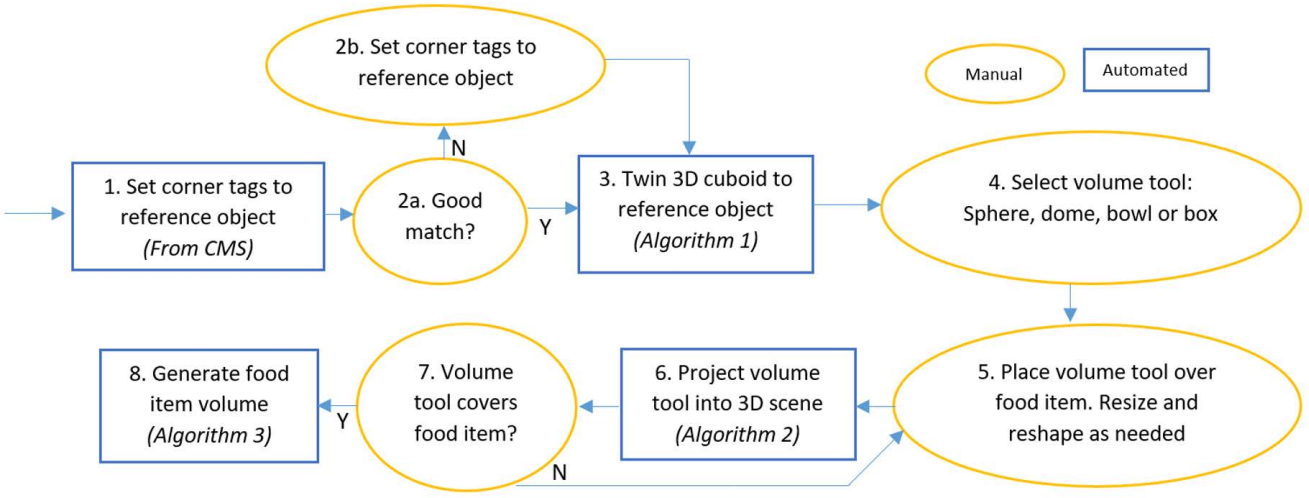

FIGURE 4. Summary of manual and automated elements of the volume estimation approach.

error rate of the new technique by comparing user trials to ground truth measures of non-food and real food items. Also the estimates provided by the users were used to determine any benefits of having multiple independent estimates, for example by allowing multiple tool users to asynchronously estimate the volumes of the same food items.

IV. EVALUATION

A. PARTICIPANTS

Participants were recruited at the University of Newcastle, Australia until 60 volunteers (22 female) completed a series of volume estimation tasks. The participants (n=60) had a mean age of 25.8 years (SD=7.1). All participants either owned or regularly used a touch-based device (for example iPhone, iPad or Samsung Galaxy Phone/Tablet). The study was approved by the University of Newcastle Human Research Ethics Committee (HREC #H-2018-0271).

B. HARDWARE

The approach was implemented as an iOS app and deployed on a 9.7' iPad. The iPad touch screen provided the interface for the manual movement of on-screen objects, for example the wire mesh objects. Standard two-finger pitch and zoom gestures were used to size and scale the wire mesh objects. The app also had a number of on-screen buttons to support the different phases of the approach (see Fig. 3). The user can also double tap the screen to hide/show the on-screen buttons as needed.

C. VOLUME ESTIMATION TASK

Participants used the volume estimation app following steps 3-8 of the process flow in Fig. 4. The corner detection was automated for all the examples in this study as we were primarily interested in the manual processes, specifically the selection, positioning and deformation of the wire mesh 3D model in steps 4-7 and how this impacted the final volume

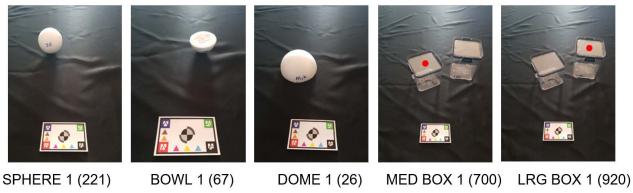

FIGURE 5. The five images used for the training trials (ground truth volume in brackets, cm^3).

accuracy. Participants completed 10 training and 16 test volume estimation trials where participants used the volume estimation technique to determine the volume of regular non-food objects and real food items in 2D images. During each training trial, after a first attempt, the participants were provided with verbal feedback on the correct volume of the target object and allowed a second try. This allowed the participants to practise with the iPad tool and gain experience with using it on 2D images. The training objects were presented as progressively more complex objects, namely styrofoam sphere, styrofoam bowl, styrofoam dome and two plastic food containers, with the latter two trials requiring both re-sizing and deformation of the virtual object (see Fig. 5). Styrofoam objects were used as their ground truth volumes were available.

After the training trials, 16 test trials were conducted with no feedback given to the participants (see Fig. 6). Participants could self-select the wire mesh shape to use from a choice of sphere, dome, bowl and cube where each default wire mesh was both resizable and deformable in all three dimensions (X, Y and Z). The ground truth volumes of the target objects was calculated by direct measurement, i.e. for the styrofoam and plastic box items, or via water displacement [4], [22] with a measuring jug measured as mL, for the food items.

Algorithm 1 Twinning 3D Cuboid to Reference Object

Input:
(1) 2D screen positions of the fiducial card's four corner tags
(2) Euler angles of 3D cuboid
(3) Increment value, set to 0.01
Output:
3D orientation of 3D cuboid twinned to the orientation of the 2D fiducial card in the image
Initialize:
(1) Boolean variables (*currentOrientation*[*.x*, *.y*, *.z*]) for increments across x, y and z Euler angles to *false*.
(2) Distance variables (*currentDistance*, *previousDistance*) between fiducial corners and 2D screen projection of 3D cuboid.

while *currentOrientation*[*.x*, *.y*, *.z*] = *false* **do** ▷ Loop until all the Euler angles across x, y and z are stable
  Increment *currentOrientation.x*
  while [LOOP.x] currentDistance >= previousDistance **do**
    increment currentOrientation.x ▷ Find best x Euler angle
  end while
  if LOOP.x triggered **then**
    currentOrientation.x = *false* ▷ Still optimising the x Euler angle so need to loop again
  else
    currentOrientation.x = *true* ▷ Best x Euler angle found for this orientation, ready to exit main loop
  end if
  Increment *currentOrientation.y*
  while [LOOP.y] currentDistance >= previousDistance **do**
    increment currentOrientation.y ▷ Find best y Euler angle
  end while
  if LOOP.y triggered **then**
    currentOrientation.y = *false* ▷ Still optimising the y Euler angle so need to loop again
  else
    currentOrientation.y = *true* ▷ Best y Euler angle found for this orientation, ready to exit main loop
  end if
  Increment *currentOrientation.z*
  while [LOOP.z] currentDistance >= previousDistance **do**
    increment currentOrientation.z ▷ Find best z Euler angle
  end while
  if LOOP.z triggered **then**
    currentOrientation.z = *false* ▷ Still optimising the z Euler angle so need to loop again
  else
    currentOrientation.z = *true* ▷ Best z Euler angle found for this orientation, ready to exit main loop
  end if
end while

Copy final 3D cuboid Euler angles to wire mesh objects (Sphere, Dome, Bowl and Box) ▷ Found the best orientation

D. PROCEDURE

Each session, of training and testing, had an average session time of 16 minutes 43 seconds (SD=21 minutes 2 seconds), and were held individually. Firstly, the research team collected signed consent forms. Secondly, the participants were given a short tutorial on the use of the new volume estimation technique as implemented on the iPad. The participants then completed the 26 volume estimation trials. Each trial had a target object highlighted with a red or white dot (white dots were used when the target food item was colored red or orange). Only one volume estimate was completed during each trial. Some images had multiple food objects and these were duplicated across the trials with the red or white dot indicating which object was the focus of the current trial.

Finally, participants completed a demographic questionnaire on gender, age, university degree or discipline area they were studying and use of touch-based devices. All participants received a $5 coffee voucher in appreciation of their participation time. Participants on eligible university courses were also given course credit as part of an assessment on research awareness activities.

Algorithm 2 Determine 3D Scene Projection of Food Item

Input:
(1) Twinned cuboid 3D corners
(2) 3D wire mesh placed over food item
Output:
(1) 3D wire mesh covering food item projected into 3D scene based on plane of twinned 3D cuboid

projectTarget = 2D screen projection of base of the 3D wire mesh
Determine plane of twinned cuboid from 3D corners
Find center point of twinned cuboid
projectPoint = 2D screen projection of 3D point.
Set *projectPoint* to cubloid center point
while *projectPoint* < *projectTarget* **do**
    Move *projectPoint* along 3D plane of cuboid by 3D unit increment
    projectPoint = 2D screen projection of 3D point
end while
Translate 3D wire mesh position to *projectTarget* ▷ wire mesh covering food item now at 3D depth of image

Algorithm 3 Volume Calculation of 3D Wire Mesh

Input:
(1) 3D wire mesh placed over food item
Output:
(1) Volume of 3D wire mesh (*theVolume*)

Determine real/virtual *ratio* by known real word reference object with scale of twinned 3D cuboid
if wire mesh = box **then** ▷ This could be 3D rectangle so need to consider each axis
    $boxWidth = ratio * geometry.boundingBox.width * wiremesh.scale.x$
    $boxHeight = ratio * geometry.boundingBox.height * wiremesh.scale.y$
    $boxLength = ratio * geometry.boundingBox.length * wiremesh.scale.z$
    $boxCubed = boxWidth * boxHeight * boxLength$
    $theVolume = boxCubed$
else if wire mesh = sphere **then** ▷ This could be an ellipoid so need to consider each axis
    $sphereRadiusX = ratio * geometry.boundingSphere.radius * wiremesh.scale.x$
    $sphereRadiusY = ratio * geometry.boundingSphere.radius * wiremesh.scale.y$
    $sphereRadiusZ = ratio * geometry.boundingSphere.radius * wiremesh.scale.z$
    $sphereCubed = sphereRadiusX * sphereRadiusY * sphereRadiusZ$
    $theVolume = (4/3) * PI * sphereCubed$
else if wire mesh = dome **then**
    $domeRadiusX = ratio * geometry.boundingDome.radius * wiremesh.scale.x$
    $domeRadiusY = ratio * geometry.boundingDome.radius * wiremesh.scale.y$
    $domeRadiusZ = ratio * geometry.boundingDome.radius * wiremesh.scale.z$
    $theVolume = ((PI * domeRadiusX * domeRadiusZ)/(3 * domeRadiusY * domeRadiusY)) * domeRadiusY * domeRadiusY * ((3 * domeRadiusY) - domeRadiusY)$
else if wire mesh = bowl **then**
    $bowlRadiusX = ratio * geometry.boundingBowl.radius * wiremesh.scale.x$
    $bowlRadiusY = ratio * geometry.boundingBowl.radius * wiremesh.scale.y$
    $bowlRadiusZ = ratio * geometry.boundingBowl.radius * wiremesh.scale.z$
    $bowlCubed = bowlRadiusX * bowlRadiusY * bowlRadiusZ$
    $theVolume = ((4/3) * PI * bowlCubed)/2$ ▷ A bowl as 1/2 sphere
end if

E. ANALYSIS

In evaluating the approach, we were interested in two primary measures, namely the volume estimation deviation from the target object's ground truth (GT) volume and the reliability of the volume judgements. The deviation was calculated by generating the absolute value (ABS) of the percentage

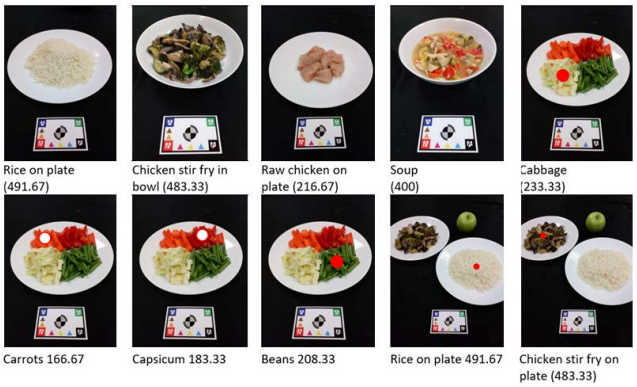

FIGURE 6. Ten of the testing images (Non-food testing objects are not shown). Food objects were a mix of food types and serving vessels (ground truth volume in brackets, cm^3).

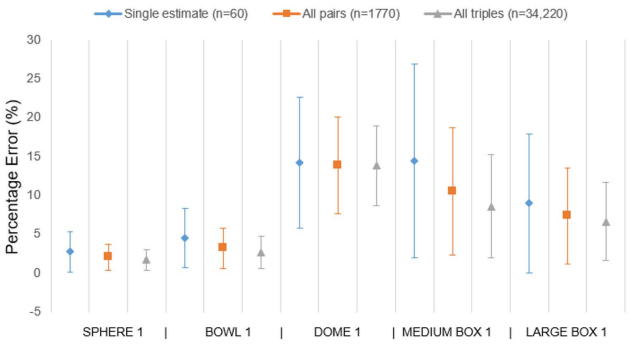

FIGURE 7. Results of the first estimate with the training images with single, pair and triple estimate combinations. Percentage error with 1 SD error bars.

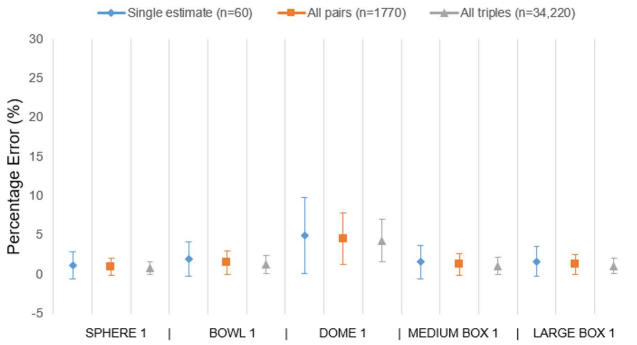

FIGURE 8. Results of the second estimate, after volume feedback, with the training images with single, pair and triple estimate combinations. Percentage error with 1 SD error bars.

deviation for each volume estimate as an absolute percentage deviation (APD):

$$APD = ABS((\frac{Estimate\_vol. - GT\_vol.}{GT\_vol.}) * 100) \quad (1)$$

Food volume estimation judgements are typically not done only once. Given the wide variation of human-based estimates [8], [23], it is common to obtain multiple estimates of food portion sizes and implicitly the volume of food items. As the current approach is a mixed method with automated and manual components, there is likely to be variation in the manual contributions. The current implementation also fits well into an overall system where crowd sourcing [23] could be used to aid scalability. Thus food items for volume estimation could be distributed to multiple app users and the results averaged in an attempt to smooth out any errors from manual components.

With a large number of food items to obtain estimates of, getting two, three or more estimates impacts further analysis of the food items, i.e. determining density and nutritional value. Thus there are two sub-questions. Firstly are there improvements in the volume estimates if more than one independent estimation is obtained (i.e. via crowd sourcing) and secondly, are there particular food types or shapes where averaging multiple independent estimates make these improvements are more prominent. Determining both these issues will highlight any return on investment on requesting multiple estimates and waiting time for multiple estimates to return.

In our study, we collected 60 estimates across 26 food items. Combinations of estimates were calculated by the number of ways to choose a sample of r unordered outcomes from a set of n possibilities:

$$C(n, r) = \binom{n}{r} = \frac{n!}{(r!(n-r)!)} \quad (2)$$

We have explored the relations between single estimates ($C(60, 1) = 60$), all pairings ($C(60, 2) = 1770$) and all triples ($C(60, 3) = 32,220$) for all 26 food item estimates. Code to generate the combinations and calculate each averaged volume estimate was written in MATLAB (version 9.7, The MathWorks Inc.).

V. RESULTS

A. TRAINING: NON-FOOD ITEMS

The results of the training estimates are shown in Figs. 7 and 8. When comparing the single estimates (n=60), it can been seen that the percentage error has been reduced in the second training trial for all volume estimates. This is to be expected as the participants were given feedback on the true volume after their first attempt. However, this is indicative that the participants are competent in the use of the iPad app for generating accuracy volume estimates as the second estimates are improved. This is particularly evident in the box objects, which required deformation of the virtual wire mesh box, where there is consistent improvement between the first and second estimates across both examples. This shows that the training was sufficient before the testing trials.

In terms of combining the estimates, both the single to paired estimates and paired to triple estimates show increased improvements and with smaller error bars. This improvement in accuracy was evident across all the testing trials.

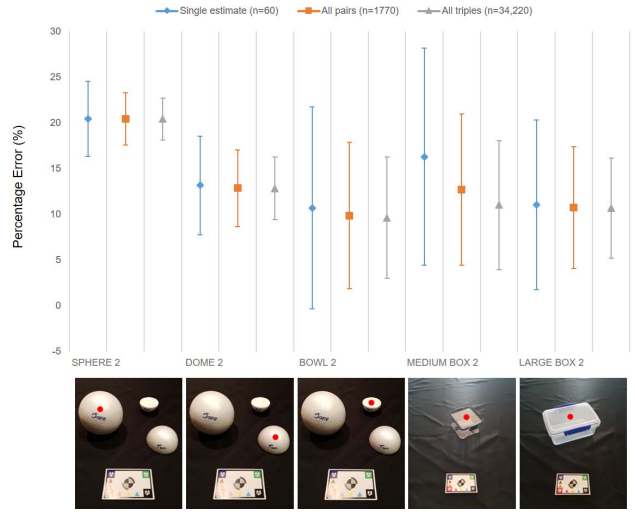

FIGURE 9. Results of the testing trials for non-food objects over mean absolute percentage deviation and percentage error.

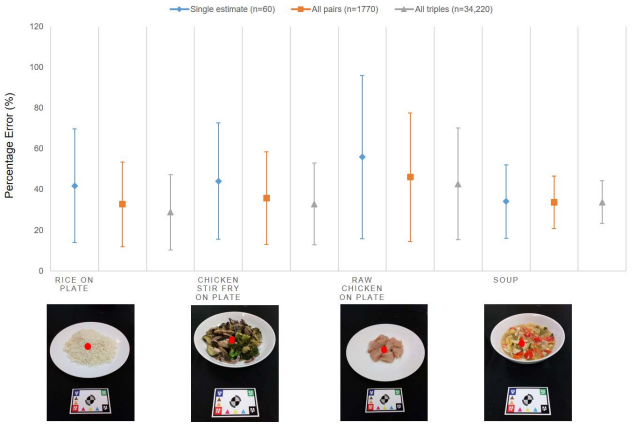

FIGURE 10. Results of the testing trials for single food items (1 food item per image) over mean absolute percentage deviation and percentage error.

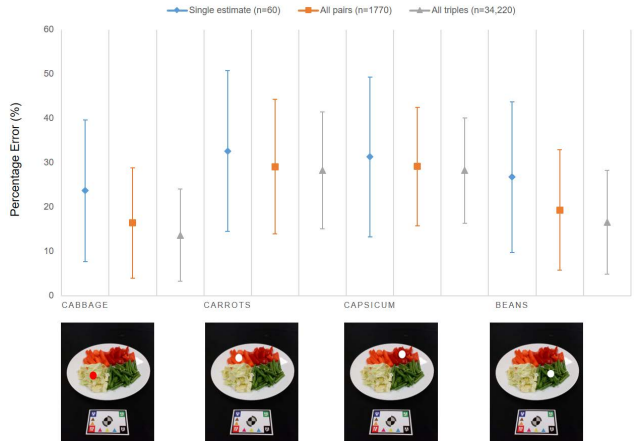

FIGURE 11. Results of the testing trials with multiple food objects on one plate. Each food item was considered individually.

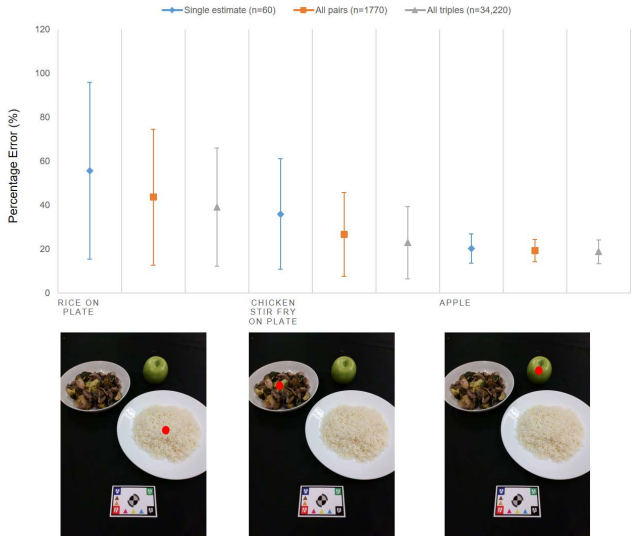

FIGURE 12. Results of testing trials with multiple food objects served discretely. Each food item was considered individually.

B. TESTING: NON-FOOD ITEMS

Fig. 9 shows the mean absolute percentage error from ground truth volume for the non-food objects under the testing conditions. Across this object set with single estimates, the worst error was 20.4% (Sphere) and the best was 10.7% (Bowl). For the multiple estimates the mean improvement was minimal, but for each increasing amount of estimates, the standard deviation and error range was reduced.

C. TESTING: SINGLE FOOD ITEMS

Fig. 10 shows the mean absolute percentage error from ground truth volume for food items where there was only one food item per image, under the testing conditions. Across this object set with single estimates, the worst error was 56.1% (Raw chicken on plate) and best was 34.2% (Soup), i.e. the more regular shaped food item gained more accurate estimates.

For the multiple estimates the mean improvement was trending towards improved, i.e. smaller, means with the best results in the triple estimate averages. Also the standard deviation and error range was reduced for each food item.

D. TESTING: MULTIPLE FOOD ITEMS

Fig. 11 shows the mean absolute percentage error from ground truth volume for multiple food objects on one plate. Across this object set with single estimates, the worst error was 32.6% (Carrots) and best was 23.7% (Cabbage). The multiple estimates had a reducing trend with increasing estimates. Also, the standard deviation and error range was reduced for each food item.

Fig. 12 shows the mean absolute percentage error from ground truth volume for multiple food objects served discretely. Across this object set with single estimates, the worst error was 55.7% (Rice on plate) and the best was 20.3% (Apple), which was expected given the more regular shape of

the apple. The multiple estimates had a reducing trend with increasing number of estimates for the irregular food items, namely the ''Rice on plate'' and ''Chicken stir fry on plate'', but less so for the ''Apple'' food item. The standard deviation and error range was reduced for each food item.

VI. DISCUSSION

The results of the current study present a method where the average error rates were in the 10-20% range for regular-shaped objects and the 10-60% range for irregular food items. Although it is difficult to directly compare this with other studies given different food item images and computational approaches that have been used previously, the results here are promising, particularly when multiple volume estimates from independent users are averaged. Also, the approach requires no prior knowledge of estimation shapes. It provides a number of common but representative proxy 3D objects, and only requires limited preparation in image capture. Further, use of the tool was surprisingly good by the participants who had no formal training with the tool, and who were from a range of discipline backgrounds. Many image based studies in a previous systematic review [15] identified that dietitians, who are highly trained in dietary assessment and portion size estimation, are not typically used to capture portion images. Thus the results of the current study provide promising insight for potentially broader use [10] and potential for real-world usage.

In terms of the potential for crowd sourcing multiple estimates to improve the reliability of the manual component of the approach, the user study provides evidence that both the median and error rates can be reduced with cross averaged pairings and triples. The generation of these combined estimate averages is trivial and the only real overhead is the generation of multiple estimates. However, with an asynchronous approach, these estimates can be collected concurrently and in the context of our larger system, distributed from a CMS to individual users on iPads. One question would be the return on investment, in time, in gaining the multiple estimates. From this study, we have seen that irregular food items gain more benefit from multiple independent estimates. One contributing factor to this it that irregular food items may have variable gaps, for example between two pieces of carrot, and this adds to estimation errors.[2] The multiple independent estimates can smooth out this gap error as the wire mesh placements under or overestimate the true volume.

For more regular shaped food items, paired estimates are sufficient. For example, in Fig. 12, the ''Rice on plate'' and ''Chicken stir fry on plate'' have improved accuracy and reducing error bars from paired to triple estimates. However, for the ''Apple'' item it would be sufficient to only have paired estimates with the sphere wire mesh shape. A similar pattern can be seen in the ''Soup'' food item in Fig. 10 where there is a good match between the regular shaped food item, i.e. soup in a bowl, and the bowl wire mesh shape.

An important aspect of the present work is that it was designed as a practical and computationally light approach that runs on a standard iPad in real-time. Based on a large user base (n=60), it exhibits high ecological fidelity [16]. The study by Fang *et al.* [7] did not focus on processing times and the required level of computation, which limits comparability to the present study. Beltran et al's [2] use of engineers who helped to develop the procedure influenced the results with Beltran *et al.* noting the dietitians had not mastered the manipulation of the wire mesh to closely conform to the outer boundary of the food image whereas the engineers who helped create the wire imaging system did.

The current model would also be appropriate to be used with other active capture systems in image based dietary assessment. Active capture methods involve the user in collecting images, for example taking the image and placing the fiducial marker in the image frame. However, it is acknowledged that this portion size estimation tool is only one aspect to dietary assessment process and for accurate real world nutrient estimation would rely on accurate food identification and up to date nutrient database information [15].

Given the results found in the current study, this approach shows the ability of the current system to be used by users with minimal training. The training used in this study comprised of only a small number of trials and transitioned well from fixed known volume objects to real foods. Interestingly some of the best estimates were found for apples and this might likely reflect the training that was conducted using spheres as these shapes have a high resemblance. Given this finding, adding additional training objects that resemble real foods might be good practise.

There was not much variation in the error estimation across the vegetables that were tested. This was expected as the vegetables were all chopped and assembled in a similar way. Raw chicken on the plate was also associated with larger errors in estimation compared to other estimations of real food. Mixed dishes or items served on shared plate present a complexity in dietary assessment as there are additional factors to consider in addition to image depth, such as the size of the overall serving vessel. This issue was highlighted in a recent review where dietary assessment from shared plates was considered [3].

VII. CONCLUSION

In the analysis of dietary intake, images of food items play a critical part of data collection. Determining the volume of these food items is necessary in order for accurate nutritional analysis. This paper has described a new semi-automated method combining 2D image capture and automated 3D scene projection with manual placement and resizing of wire mesh objects to generate volume estimates of food items. Through a combination of automated and manual activities, the volumes of food items in 2D images can be generated in a

[2]For food items with natural gaps, for example rice or pasta, if there is an associated measures database there may be measures that relate to volume with a weight equivalent. For other items, the coverage of the food by the 3D wire mesh is determined by the analyst/tool user. Therefore multiple estimates are desirable for irregular food items.

process that is computationally light and runs in real-time. Also we have considered the use of multiple food volume estimates and showed that accuracy can be increased by combining multiple independent estimates.

However, the approach presented is not without limitations. There can be errors in both the automated and manual elements and, as a pipeline approach, any inaccuracies accumulate and increase the final error [20]. Thus, reducing any error is desirable. Also with increasing availability of wearable sensors (see [19]) there are opportunities for hybrid approaches combining direct (based on physical properties of food) and indirect (based on food intake activity) measures to reduce overall error. Future work will focus on improving the accuracy of the automated aspects of the approach, for example the accurate identification of corner markers that is critical to accurate 3D projection, and determining the impact of increased training [12] for improvements in the manual aspects, i.e. the fitting and sizing of the wire mesh objects.

REFERENCES

[1] B. Amoutzopoulos, P. Page, C. Roberts, M. Roe, J. Cade, T. Steer, R. Baker, T. Hawes, C. Galloway, D. Yu, and E. Almiron-Roig, "Portion size estimation in dietary assessment: A systematic review of existing tools, their strengths and limitations," *Nutrition Rev.*, vol. 78, no. 11, pp. 885–900, Nov. 2020, doi: 10.1093/nutrit/nuz107.

[2] A. Beltran, H. Dadabhoy, C. Ryan, R. Dholakia, J. Baranowski, Y. Li, G. Yan, W. Jia, M. Sun, and T. Baranowski, "Reliability and validity of food portion size estimation from images using manual flexible digital virtual meshes," *Public Health Nutrition*, vol. 22, no. 7, pp. 1153–1159, 2019, doi: 10.1017/S1368980017004293.

[3] T. Burrows, C. Collins, M. Adam, K. Duncanson, and M. Rollo, "Dietary assessment of shared plate eating: A missing link," *Nutrients*, vol. 11, no. 4, p. 789, Apr. 2019, doi: 10.3390/nu11040789.

[4] A. Concha-Meyer, J. Eifert, H. Wang, and G. Sanglay, "Volume estimation of strawberries, mushrooms, and tomatoes with a machine vision system," *Int. J. Food Properties*, vol. 21, no. 1, pp. 1867–1874, Jan. 2018, doi: 10.1080/10942912.2018.1508156.

[5] N. M. de Vlieger, M. Weltert, A. Molenaar, T. A. McCaffrey, M. E. Rollo, H. Truby, B. Livingstone, S. I. Kirkpatrick, C. J. Boushey, D. A. Kerr, C. E. Collins, and T. Bucher, "A systematic review of recall errors associated with portion size estimation aids in children," *Appetite*, vol. 147, Apr. 2020, Art. no. 104522, doi: 10.1016/j.appet.2019.104522.

[6] J. Dehais, M. Anthimopoulos, S. Shevchik, and S. Mougiakakou, "Two-view 3D reconstruction for food volume estimation," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1090–1099, May 2017, doi: 10.1109/TMM.2016.2642792.

[7] S. Fang, C. Liu, F. Zhu, E. J. Delp, and C. J. Boushey, "Single-view food portion estimation based on geometric models," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 385–390, doi: 10.1109/ISM.2015.67.

[8] G. P. Faulkner, M. B. E. Livingstone, L. K. Pourshahidi, M. Spence, M. Dean, S. O'Brien, E. R. Gibney, J. M. Wallace, T. A. McCaffrey, and M. A. Kerr, "An evaluation of portion size estimation aids: Precision, ease of use and likelihood of future use," *Public Health Nutrition*, vol. 19, no. 13, pp. 2377–2387, Sep. 2016, doi: 10.1017/S1368980016000082.

[9] C. Frobisher and S. M. Maxwell, "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. Nutrition Dietetics*, vol. 16, no. 3, pp. 181–188, 2003, doi: 10.1046/j.1365-277X.2003.00434.x.

[10] D. K. N. Ho, W.-C. Chiu, Y.-C. Lee, H.-Y. Su, C.-C. Chang, C.-Y. Yao, K.-L. Hua, H.-K. Chu, C.-Y. Hsu, and J.-S. Chang, "Integration of an image-based dietary assessment paradigm into dietetic training improves food portion estimates by future dietitians," *Nutrients*, vol. 13, no. 1, p. 175, Jan. 2021, doi: 10.3390/nu13010175.

[11] D. K. N. Ho, S.-H. Tseng, M.-C. Wu, C.-K. Shih, A. P. Atika, Y.-C. Chen, and J.-S. Chang, "Validity of image-based dietary assessment methods: A systematic review and meta-analysis," *Clin. Nutrition*, vol. 39, no. 10, pp. 2945–2959, Oct. 2020, doi: 10.1016/j.clnu.2020.08.002.

[12] A. Hooper, A. McMahon, and Y. Probst, "The role of various forms of training on improved accuracy of food-portion estimation skills: A systematic review of the literature,," *Adv. Nutrition*, vol. 10, no. 1, pp. 43–50, Jan. 2019, doi: 10.1093/advances/nmy060.

[13] C. D. Lee, J. Chae, T. E. Schap, D. A. Kerr, E. J. Delp, D. S. Ebert, and C. J. Boushey, "Comparison of known food weights with image-based portion-size automated estimation and Adolescents' self-reported portion size," *J. Diabetes Sci. Technol.*, vol. 6, no. 2, pp. 428–434, Mar. 2012, doi: 10.1177/193229681200600231.

[14] Y. Liu, J. Lai, W. Sun, Z. Wei, A. Liu, W. Gong, and Y. Yang, "Food volume estimation based on reference," in *Proc. 4th Int. Conf. Innov. Artif. Intell.*, May 2020, pp. 84–89, doi: 10.1145/3390557.3394123.

[15] F. P. W. Lo, Y. Sun, J. Qiu, and B. Lo, "Image-based food classification and volume estimation for dietary assessment: A review," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1926–1939, Jul. 2020, doi: 10.1109/JBHI.2020.2987943.

[16] R. P. McMahan, E. D. Ragan, A. Leal, R. J. Beaton, and D. A. Bowman, "Considerations for the use of commercial video games in controlled experiments," *Entertainment Comput.*, vol. 2, no. 1, pp. 3–9, Jan. 2011, doi: 10.1016/j.entcom.2011.03.002.

[17] K. Okamoto and K. Yanai, "An automatic calorie estimation system of food images on a smartphone," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manage.*, New York, NY, USA: Association for Computing Machinery, 2016, pp. 63–70, doi: 10.1145/2986035.2986040.

[18] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–8, doi: 10.1109/WACV.2009.5403087.

[19] V. B. Raju and E. Sazonov, "A systematic review of sensor-based methodologies for food portion size estimation," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12882–12899, Jun. 2021, doi: 10.1109/JSEN.2020.3041023.

[20] M. E. Rollo, T. Bucher, S. P. Smith, and C. E. Collins, "ServAR: An augmented reality tool to guide the serving of food," *Int. J. Behav. Nutrition Phys. Activity*, vol. 14, no. 1, pp. 1–10, Dec. 2017, doi: 10.1186/s12966-017-0516-9.

[21] G. A. Tahir and C. K. Loo, "A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment," *Healthcare*, vol. 9, no. 12, p. 1676, Dec. 2021, doi: 10.3390/healthcare9121676.

[22] Y. Yang, W. Jia, T. Bucher, H. Zhang, and M. Sun, "Image-based food portion size estimation using a smartphone without a fiducial marker," *Public Health Nutrition*, vol. 22, no. 7, pp. 1180–1192, May 2019, doi: 10.1017/S136898001800054X.

[23] J. Zhou, D. Bell, S. Nusrat, M. Hingle, M. Surdeanu, and S. Kobourov, "Calorie estimation from pictures of food: Crowdsourcing study," *Interact. J. Med. Res.*, vol. 7, no. 2, p. e17, Nov. 2018, doi: 10.2196/ijmr.9359.

SHAMUS P. SMITH (Senior Member, IEEE) received the B.Sc., B.Sc. (Hons.), and Ph.D. degrees in computer science from Massey University, New Zealand, in 1992, 1993, and 1999, respectively. He is currently an Associate Professor in Computing and IT at The University of Newcastle, Australia. He has published over 100 research articles. His research interests include virtual reality, human–computer interaction, and technology enhanced learning.

MARC T. P. ADAM received the bachelor's degree in computer science from the University of Applied Sciences Würzburg, Germany, and the Ph.D. degree in information systems from the Karlsruhe Institute of Technology, Germany. He is currently an Associate Professor in Computing and IT at The University of Newcastle, Australia. His research interests include interplay of user cognition and affect in human–computer interaction. He is a Founding Member of the Society for NeuroIS.

GRACE MANNING is currently an Accredited Practicing Dietitian, a Graphic Designer, and a Research Assistant at the Priority Research Centre for Physical Activity and Nutrition, School of Health Sciences, The University of Newcastle, Australia. Her work spans a range of nutrition areas, including aged care, nutrition in pregnancy, cooking, disability, nutrition education, and technology. Her research interests include communication and marketing and environmental sustainability.

TRACY BURROWS is currently a Professor in Nutrition and Dietetics at The University of Newcastle, Australia. She is also a researcher at the Hunter Medical Research Institute. Her research interests include dietary assessment, eating behaviors, in addition to the management of overweight, obesity, and addictive eating. She is a National Health and Medical Research Fellow.

CLARE COLLINS is currently a Laureate Professor in Nutrition and Dietetics and the Director of Research at the School of Health Sciences, The University of Newcastle, Australia. Her research interests include personalized food and nutrition eHealth programs, tools and evaluating impact on eating patterns and diet-related health across key life stages, and chronic disease conditions.

MEGAN E. ROLLO received the B.App.Sci., B.Hlth.Sci. (Nutrition and Diet), and Ph.D. degrees from the Queensland University of Technology, Australia. She is currently a Research Fellow in Nutrition and Dietetics with the School of Health Sciences and the Priority Research Centre for Physical Activity and Nutrition, The University of Newcastle, Australia. Her research interests include technology-assisted dietary assessment and personalized behavioral nutrition interventions.