A Systematic Review of Sensor-Based Methodologies for Food Portion Size Estimation

Viprav B. Raju*, Member, IEEE, and Edward Sazonov*, Senior Member, IEEE

Abstract—Food portion size estimation (FPSE) is critical in dietary assessment and energy intake estimation. Traditional methods such as visual estimation are now replaced by faster, more accurate sensor-based methods. This article presents a comprehensive review of the use of sensor methodologies for portion size estimation. The review was conducted using the PRISMA guidelines and full texts of 67 scientific articles were reviewed. The contributions of this article are threefold: i) A taxonomy for sensor-based (SB) FPSE methods was identified, classifying the sensors (as wearable, portable and stationary) and the methodology (as direct and indirect). ii) A novel comprehensive review of the state-of-the-art SB-FPSE methods was conducted and 5 sensor modalities (Acoustic, Strain, Imaging, Weighing, and Motion sensors) were identified. iii) The accuracy of portion size estimation and the applicability to free-living conditions of these SB-FPSE methods were assessed. This article concludes with a discussion of challenges and future trends of SB-FPSE.

Index Terms—Sensors, food intake sensors, food portion size estimation, food volume, dietary assessment.

I. INTRODUCTION

Dietary intake assessment has been one of the profound areas of scientific research, particularly in assessing the impact of diet on the human body and an individual's state of well-being [1], [2]. The World Health Organization (WHO) underlines the importance of a healthy diet and contends that unhealthy eating habits, along with the lack of physical exercise, manifests into global health risks [3]. Establishing a desirable diet is dependent on each individual. Determining dietary guidelines is not only pertinent on an individual basis but also has global implications [4]. It is essential to identify the dietary restrictions that should be followed to lead a healthy life and make additions to the diet to have a balanced energy intake.

Energy density and portion size have been identified as properties of foods that can modulate energy intake [5]. Energy density refers to the amount of energy per unit mass of food (kcal/g). Portion sizes can be expressed as mass (g), volume (ml), household measures (e.g. tablespoons), hand measures (e.g. a fist), or as measures relative to the size of a reference object (e.g. “tennis ball” [6]).

A large portion size increases energy intake in individuals [5], [7], [8]. This is known as the ‘Portion Size Effect’ [9]. Dietary interventions such as weight control rely on an individual’s ability to choose appropriate portion sizes and that person’s awareness about the amount of food being consumed [10]. Therefore, it is critical to develop accurate methods for food portion size estimation (FPSE).

FPSE methods can be broadly classified into traditional-FPSE (T-FPSE) and SB-FPSE. In T-FPSE, portion sizes of the food consumed are estimated by directly measuring the quantity of food (mass/volume) using household measures or by visual approximation. Household measures such as calibrated measuring cups/jars or spoons can be used to measure the quantity of food directly. This can be quite cumbersome and it is not possible to directly measure the food quantity for foods consumed outside of the household. Visual guides are frequently used to estimate portion sizes. Hands are often used as a visual guide by healthcare professionals and the common public to estimate portion size as described in Fig. 1. For example, a fist, a thumb fingertip, and an index fingertip are used to estimate one cup, one tablespoon, and one teaspoon, respectively. The accuracy of such estimates is very low. Visual estimates are used in two popular reporting methods: 24-hour recall [11]–[14] and food diaries [15].

There are numerous drawbacks in T-FPSE and related reporting methods. It was seen that individuals have difficulty assessing food portion sizes [6], [11], [12] and are...
confused as measurement units/terminologies are used inconsistently [13]. The accuracy of T-FPSE varies in reporting methods that involve conceptualization and memory [14], [15]. Traditional methods suffer from a flat-slope phenomenon, in which large portions tend to be underestimated and small ones overestimated [14], [16], [17]. Amorphous foods (e.g. mashed potatoes) and those eaten in small portions are reported less accurately than other types of foods [10], [16], [18]. Also, T-FPSE based reporting methods are cumbersome, and prone to errors due to memory, particularly in recall methods. The intake behavior may also be affected in these methods. The academic community, since, has extensively explored the use of technological aids such as sensors for FPSE.

Sensor-based technology is gaining popularity in dietary assessment. Sensors have been used to address a variety of problems, such as automatic food intake detection (utilizing piezoelectric sensors [19], capacitive sensors [20], accelerometers [21], and others), food type recognition, energy intake estimation [22], and food portion size estimation. Several existing reviews covered various uses of sensors in food intake [23], [24], but presently there are no reviews focused on the sensor-based portion size estimation.

The present review is intended to provide a systematic, comprehensive evaluation of state-of-the-art sensor-based methods for FPSE. To cover the full range of approaches in this survey, research publications were thoroughly studied, and a total of 121 papers (without duplication) were found related to the topics. Following the application of inclusion and exclusion criteria, 67 papers were selected for full-text review.

The contributions of this review are three-fold: i) A taxonomy for SB-FPSE methods was identified ii) A novel comprehensive review of the state-of-the-art SB-FPSE methods was conducted iii) The accuracy of portion size estimation and the applicability to free-living conditions of these SB-FPSE methods were assessed.

The paper is organized as follows. First, the methodology of the systematic review is presented in Section II along with the specification of research questions (RQs). The review findings, which include taxonomy, are discussed in Section III. Section III also presents a detailed description of the articles included in the review categorized by the type of sensor modality. Sections IV, V, and VI are discussion, open problems/future directions, and conclusion, respectively.

II. REVIEW METHODOLOGY

Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) [25] was used as a guide for this review. The authors independently screened the titles and abstracts of the publications retrieved through the database search and then carried out a full-text review of all relevant studies. This methodology used the following processes:

A. Identifying Research Questions

Three research questions were chosen to guide this systematic review:

RQ1) What are the available state-of-the-art SB-FPSE methodologies?

The answer to this question helps in identifying the sensor modalities that have been used for portion size estimation. The authors independently screened the titles and abstracts of the publications retrieved through the database search and then carried out a full-text review of all relevant studies. This methodology used the following processes:

RQ2) What methods are employed for portion size estimation from sensor data and how accurate are these methods?

The answer to this question enables the authors to assess and compare the validity of various sensor-based methods.

RQ3) Which sensor modalities are more suitable for use in the free-living conditions?

The answer to this question helps in determining the applicability of the sensor-based methods to studies outside of the lab and identifying the research gaps in current methodologies.
Exhustive electronic searches for relevant literature were performed across five repositories: PUBMED, Science Direct, SCOPUS, ACM Digital library, and IEEE Explore from inception through May 21st, 2020.

C. Search Strategy

To cover all existing approaches for portion size estimation, the following keywords were considered: ‘sensor’, ‘technology’, ‘device’, ‘food portion size’, ‘food mass estimation’, and ‘food volume’.

The query string used for the search was: “Sensor” OR “Device” OR “Technology”) AND (“Food Portion Size” OR “Food Volume” OR “Food Mass Estimation”)

The search results were restricted to the English language. References from the selected primary full-text articles were further analyzed for relevant publications. The selection was further narrowed by applying the eligibility criteria described in Table I. The manual bibliographic search identified articles through sources outside the mentioned databases such as reference lists of articles. Cited references or citing references of articles included through the database search were examined, and suitable ones were added to the survey.

Articles that fulfilled the inclusion criteria were considered in this review and those that fulfilled the exclusion criteria were filtered out.

D. Results

Initial electronic database searches resulted in a total of 2368 publications. A manual bibliographic search also identified 6 additional publications that were qualified to be included. A total of 2215 articles were set aside for screening based on abstracts and titles, of these, 121 articles were selected for full-text review. Following the exclusion process, 67 articles were included in the systematic review based on the eligibility criteria. Fig. 2 illustrates the methodology and results of the review process.
III. Review Findings

A. Taxonomy

First, the taxonomy for SB-FPSE was identified to address RQ1. A potential classification can be done based on the type of instrumented devices: Wearable, Portable, and Stationary devices. Wearable devices, as the name implies, are integrated into wearable objects or directly with the body [26]. Portable devices are those that may be carried around, such as a smartphone. Stationary sensors are those sensors that are embedded into the environment and stationary objects. Fig. 3 describes a taxonomy that was established from the findings of this review.

Another classification scheme is based on the measurand: Direct and Indirect methods of FPSE. Direct methods estimate portion size by measuring the properties of food such as food mass or food volume directly. Indirect methods are those that use behavioral and physiological manifestations of eating such as hand gestures, bites, chews, and swallows to approximate the ingested amount.

B. Sensor Modalities

The review identified 5 sensor modalities that were primarily used for FPSE. They were: Acoustic, Strain, Imaging, Weighing, and Motion sensors. One immediate observation is that the subset of sensors used for FPSE is much narrower than the sensors used, for example, for the detection of food intake. Fig. 4 provides categories of the sensors included in this review.

Acoustic, strain, and motion sensors were involved in the indirect measurement of portion size and were based on food intake activities such as chewing, swallowing, hand gestures, and head movement. All the sensors used in indirect FPSE were found to be wearable sensors.

Weighing and imaging sensors were used for more direct methods of FPSE that were based on the physical properties of the food item such as mass and volume. Imaging sensors were either wearable, portable, or stationary while weighing sensors were either stationary or portable. Each modality is described in detail below.

C. Acoustic Sensors

Acoustic wearable sensors have been previously used to monitor food intake [27]. The authors suggested two possible locations for an acoustic wearable sensor (Fig. 5), each associated with a physiological activity: chewing and swallowing. A few recent studies have employed these sensors in FPSE. All these studies used methods that can be grouped under indirect measurement of portion size.

One study [28] presented an approach to predict the weight of individual bites taken using an ear-pad chewing acoustic sensor. Chewing sounds were recorded using a miniature microphone (Knowles, TM-24546) embedded in a custom ear-pad device. A pattern recognition algorithm was developed to identify chewing cycles and to detect the food type. The bite weights were predicted based on this information.

The study included 8 participants and 504 habitual bites. The acoustic sensor system was used to predict bite weight for three types of foods (potato chips, lettuce, and apple) using linear regression models with prediction errors in bite weights from 19.4% to 31%. The authors claimed that the bite weight prediction using acoustic chewing recordings is a feasible approach for solid foods and should be further investigated.

Researchers have claimed that the information extracted from the temporal sequence of chews and swallows can be used to estimate the mass of food consumed [29]. The counts of chews and swallows along with other derived metrics were used to build prediction models for detection of food intake, differentiation between liquids and solids, and estimation of the mass of ingested food. The proposed prediction models were able to predict mass of ingested food with an accuracy of 91% for a limited number of solids and an accuracy of 83% for liquids. The insight obtained from that preliminary study on mass intake estimation was used to design a new study [30]. The study individually calibrated parameters that were computed for each subject based on the total number of chews and swallows observed in the training meals. The average mass per chew of solid foods, the average mass per swallow of solid foods, and the average mass per swallow of liquids were computed. These were used to create models to predict the mass of consumed solids and liquids. Counts of chews and swallows were used to estimate the mass of individual food items with the mean absolute percentage error of 32.2% ± 24.8%.
The authors in [31] proposed the use of multi-modal wearable sensors, such as an in-ear acoustic sensor plus head and wrist motion sensors. The authors stated that knowing the food type in combination with sensor data was critical for estimating the food amount consumed. A customized earbud and a pocket audio recorder were used as internal and external microphones. The external microphone was used to remove external noise while the subtle eating signal was captured mainly on the internal microphone. After noise cancellation, each intake window was split into frames. The frame length was chosen to capture a full chew without including multiple chews in a single frame. For each frame, the following features were computed: energy, spectral flux, zero-crossing rate, and 11 Mel-frequency cepstral coefficients (MFCCs). Then, the mean and standard deviation of the frame features were stored as a feature vector for the whole intake window. A random forest regression with 40 trees was used that included all features (audio, motion, and annotation). Data was collected with people wearing acoustic and motion sensors, with ground truth annotated from video and continuous scale data. The error in weight estimation was 35.4%.

D. Strain Sensors

Wearable strain sensors have been used for indirect FPSE by sensing chewing. The authors in [30] used a strain sensor along with an acoustic sensor to demonstrate the suitability of estimating portion size using individually-calibrated models based on Counts of Chews and Swallows (CCS models). The study was conducted under laboratory settings. Chew counts and swallow counts were estimated using sensor signals and video recordings. CCS models were compared to diet diaries and it was found that CCS models presented the lowest reporting bias and a lower error as compared to diet diaries. They achieved a mean absolute percentage error of 32.2% ± 24.8% while estimating portion sizes.

The same research group conducted another study using just the strain sensor [32]. In both studies, a piezoelectric strain sensor (LDT0-028K, Meas-Spec Inc.) was attached to the skin immediately below the earlobe using a medical adhesive (Hollister 7730). The strain sensor (Fig. 7) allowed the monitoring of jaw motion during chewing. Subject-independent models were derived from bite, chew, and swallow features obtained from either video observation or information extracted from the sensor. 30 participants each consumed 4 different meals in a laboratory setting while being videotaped. Research staff kept a record of mass of each bite by recording the food weight, before and after each bite, using a commercially available scale with 1 g precision. With multiple regression analysis, a forward selection procedure was used to choose the best feature set. The best estimate of meal mass had an absolute percentage error of 25.2% ± 18.9% (mean ± standard deviation), and estimation errors of −17.7 ± 226.9 g (mean ± standard deviation).

E. Imaging Sensors

Imaging sensor-based methods for portion size can be grouped under direct FPSE and may fall under wearable, portable or stationary categories. Also varying are types of imaging devices, number of viewpoints, use of fiducial markers, passive/active capture of images (whether the image is taken by the user or automatically), and others. We have summarized the studies included under imaging sensors in Tables II, III, and IV (see Appendix A) which include portable, stationary, and wearable imaging sensors, respectively.

The types of devices used for image capture include video cameras, handicams, smartphones, DSLRs, and custom-made cameras. The type of camera (imaging device) and the image resolution are very important in food photography [33], [34]. A clear food image provides a better input source to both humans and computers alike.

The number of viewpoints may be classified as either single-view or multi-view. The review identified that there were essentially two categories in imaging namely single-view
color/color-depth imaging and multi-view color imaging. Color imaging involves recording color information of the scene of interest using imaging photosensors. Color filters are used along with the photosensors to filter the light by wavelength range, such that the separate filtered intensities include information about the color of light. The recorded information is then used to reproduce the original colors by mixing various proportions of the captured color channels (example: Red-Green-Blue (RGB)). Depth sensing captures the depth information as an additional stream of information.

The fiducial markers may be used as a dimensional and color reference. Fiducial markers are used as a point of reference or measure and are placed in the field of view of the imaging system. Volume estimation using imaging needs a dimensional reference whose volume or dimensions are known. This dimensional reference may include: custom made backgrounds [35]; circular references (plates, bowls, or food containers) [36]; colored checkerboard [37]; business cards [38], [39]; the international food unit (IFU) [40], [41]; laser mesh or pattern [42], [43]. The type of visual cue used determines the complexity of the setup. Some methods require the user to carry around the reference (checkerboards, blocks, or cards).

The passive or active mode of capture defines the need for a user to take action during each eating episode, however small. “Passive” methods use wearable sensors to trigger the camera [44]–[47], not requiring any action by the user. “Active” methods require the user to capture the food images during every eating episode, thus making the result dependent on the quality of self-report.

The number of viewpoints was determined as the most important criterion, and the review results were grouped either as single-view or multi-view methods.

1) Single-View Methods: The single-view color/color+depth approach is defined as any method where images of food are captured before, during or after the eating episode by a single image sensor and at a one-time instance. For example, a user may use a smartphone to take an image of the serving size before a meal, and an image of the leftovers after the meal.

Since there is only a single viewpoint, there is a need for a dimensional reference such as a fiducial marker or any other visual cues in the image scene. The review identified that the use of such references was more prevalent in these methods. Several studies [48]–[50] have used fiducial markers that need to be carried around by users which often induce user burden. Researchers also evaluated other dimensional references that are naturally found in the image scene and do not require additional fiducial markers. One example is the use of cutlery or chopsticks as a dimensional reference [51], [52]. One study that involved portable imaging sensors used such a reference object, particularly participants’ thumbs as a reference for FPS estimates [53]. The authors used the known size of the thumb and estimated the portion size by comparing the food and thumb in the images. Other popular reference objects are food containers such as plates or bowls. The method reported in [54] used food shape models using plates and bowls as references; achieving error rates of 3.7%, 19.65%, and 15.36% for a wafer box, chips, and cheese, respectively, while measuring volume. This method assumes the presence of a plate that is easy to identify, however, many foods are eaten without plates. Authors in [55] used templates of food shapes. In this approach, food images are segmented and assigned a food name/code. A pre-defined system chooses a food template shape for each segmented food item. The authors reported an error of 11% for beverages and an 8% error for solids (bread slices) in volume estimation. A similar study [56] calculated the food location in the 3-dimensional coordinate system using the plate as a scale reference. After segmentation and further image processing, a shape was picked from a pre-constructed shape library. The authors reported an average error rate of 3.9% in food volume.

As opposed to the use of physical dimensional references, some researchers explored the use of virtual 3D wireframes/meshes. A virtual 3D wireframe is projected into the food image in a proportional relationship. Using this relationship, the known volume of the wireframe is utilized to compute the food volume. This method was robust in estimating the FPS under several different experimental conditions and gave identical results for different cameras [57]. The study involving a portable sensor [57] reported a relative error in volume estimation around 7% for regularly shaped food and around 19% for irregularly shaped food. The study involved 2 regularly shaped objects and 19 food replicas. Another research group that used a portable sensor reported an average error rate in volume of 10% in 224 trials, consisting of a cuboid and 7 NASCO food replicas [58]. Virtual meshes/wireframes have also been used in studies involving wearable sensors. One research group used a wearable sensor and reported an average error rate of $-2.8 \pm 20 \%$ in the volume estimation of 50 samples of Western/Asian foods [59]. The authors in [60], [61] discussed the validity of the wireframe procedure using a wearable camera. However, wireframes can be erroneous for irregular shaped food and might be tedious due to necessary manual inputs. These studies involving wearable cameras performed well during portion size estimation which hints at the possibility of accurate FPS estimates using wireframes for the data captured in free-living. A different type of virtual input was proposed and evaluated by researchers using a portable imaging sensor [62]. The authors introduce an International Food Unit (IFU). The approach relies on a virtual cube to relate the food size with the cube size. They achieved error rates of less than 20% for large volume foods and around 40% for small volume foods while measuring volume.

Another popular approach in single-view methods is the use of geometric models. The authors in [38] reported that FPSE using geometric models such as cylinders, prisms, and spheres produced better results in terms of accuracy of estimation when compared to depth images captured by structured light scanners. Irregular shaped food items may be a challenge to the approach. In another study involving geometric models [51], the authors reported an average error rate of 6% in energy estimation. The accuracy of FPSE was further refined in a following study that made use of food co-occurrence patterns. Food portion co-occurrence pattern is a type of contextual information where patterns from food images are used to study
the user’s eating behavior [39]. Both of these studies [38], [51] employed portable imaging sensors.

Recent developments in sensor technology and methodology have enabled researchers to evaluate latest sensors such as off-the-shelf 3D scanners [63]–[66] and latest methodology such as deep learning [67]–[71] in FPSE. Another study used short-range depth cameras to determine FPS [72]. The authors achieved a low error rate of 7.5% in volume estimation and suggested that the use of these off-the-shelf 3D scanners could be a possible method for accurate FPSE. However, these scanners can be expensive and could be difficult to use routinely. A relatively recent method in single-view FPSE is the use of neural-nets. Im2Calories proposed in [73] makes use of a Convolutional Neural Network (CNN) [74]. This approach using a portable sensor was accurate in EI estimates and worked well for combinations of food items.

Stationary imaging sensors are fixed to a system that is not intended to move. The system is usually located in a laboratory. Single-view methods using stationary imaging sensors have been employed previously and the accuracy in portion size of these systems were found to be higher compared to other sensor types. One study estimated the volume of food using the intrinsic properties of the camera such as resolution and focal length [75]. The authors achieved an error rate of 13.3% in volume estimation. Another group used cubic spline interpolation [76] and achieved an error rate of 0.625% in volume estimation of six synthetic objects. A stationary depth sensor was used in another study [77], to achieve an error rate of 3% in volume estimation.

2) Multi-View Methods: A multi-view method is any method where images of food are captured before, during or after the eating episode by either a single image sensor at multiple view angles or by multiple imaging sensors.

The review identified that multi-view methods have been less explored compared to single-view methods but are gaining popularity. The advantage of having more than one viewpoint enables the sensor system to mimic the human eye (binocular vision-based depth perception). The need for a dimensional reference is frequently eliminated as there are at least two different perspectives of the same image scene.

One of the popular multi-view approaches is the stereo-imaging-based 3D reconstruction [78]–[81]. A study involving a portable sensor [78] explored the stereo reconstruction approach by considering real food datasets with more than one food item. The authors claimed that the stereo approach works well for FPSE but might have issues in the case of textureless food items and non-uniform lighting conditions. The average error in volume was 5.75 (± 3.75)%.

Binocular vision-based methods have also been explored using wearable cameras. The authors in [82] used a wearable dual-lens camera system on 16 types of food objects with error rates of 3.01% in food dimensions and 4.48% in food volume. The study of [83] used a wearable stereo camera and achieved an error rate of 2.4% in volume estimation. Both approaches [82], [83] worked well for irregular shaped food items. However, two-view stereo wearable devices were not tested in free-living.

Some studies have explored the idea of using dimensional references, like in single-view methods. The reference object along with the food is photographed in two different perspectives, for example a top-view and a side-view. The dimensions (length, width, height) of the food item are calculated first, using the reference object, and then the volume is calculated. The references may include thumbs, hands or random objects [84]–[88].

Some multi-view methodologies use more than two viewpoints. These methods have been used less frequently in the FPSE. Increasing the number of views results in increased complexity of the sensor system. Shape from silhouettes is a popular 3D reconstruction approach [89], [90] that estimates a 3D shape estimate of an object using silhouette images of the object. This concept has been applied in FPSE to generate 3D volumes of food items [91], [92] using a portable sensor. The camera intrinsic and extrinsic parameters are determined for every image and each estimation process needs several images taken from different points of view. However, compared to other methods, it appears to be robust to segmentation noise. The average error in volume estimation was 9.5-10%. Furthermore, this approach does not require any prior shape information from food identification, and it may work for arbitrarily shaped food objects. This method might be cumbersome and more suited to laboratory settings. The applicability to free-living conditions can be questionable.

Another study [93] discussed a technique that estimates the volume directly from three orthogonal images without explicit 3D reconstruction of the object’s shape. The proposed approach makes no assumptions about the shape of the object, except that it is convex, axially symmetric, and axially aligned. They achieved an average FPSE error rate of 11.9% for 8 real foods consisting of vegetables, fruits, and 8 artificial geometric 3D objects.

A recent study involving a portable and a wearable imaging sensor [46] explored the idea of simultaneous localization and mapping (SLAM). Sparse maps from SLAM was generated and further processed using a multiple convex hull algorithm to acquire a mesh object. They achieved an overall error of 17% in volume estimation. This approach doesn’t require any prior information on the food shape. Another approach, similar to SLAM, is the structure from motion algorithm (SFM) [94]. This approach produced accurate results for convex food items using a portable sensor. The algorithm first segmented the images into food items and extracted the volume by applying the SFM algorithm on 6 separate frames. The accuracy of volume estimation (ml) was 92% for real single food items. Another study [67] proposed a multi-view structured light 3D reconstruction technique. A food model was synthesized using multiple depth maps obtained from multiple image reconstructions.

F. Motion Sensors

Motion sensors include accelerometers, gyroscopes, and Inertial Measurement Units (IMU). These sensors have been used to track characteristic hand-to-mouth gestures, demonstrating a pattern of wrist-roll while taking a bite (Fig. 8) [95], [96].
However, this review identified only one study that employed wearable motion sensors for indirect estimation of portion size. The study of [31] used a multi-sensor system (also discussed under acoustic sensors). Authors utilized features derived from acoustic sensors (zero-crossing rate, energy, spectral flux, etc.), motion sensors (zero-crossing rates, temporal shape features, etc.), and food type to estimate the mass of food intake. The mean absolute percentage error in food weight estimation was 35.4%.

G. Weighing Sensors

Weighing is a direct method of measuring the food quantity where the mass of the food item is measured. Weighing sensors consist of load cells that are embedded in devices such as a weigh-scale. Recently, these sensors have been used in FPSE using the motivation that people generally eat on a solid surface where these sensors can be positioned. There are two types of weighing sensors: stationary and portable weighing sensors. We have summarized the studies included under weighing sensors in Table V (see Appendix A).

Stationary weighing sensors are generally part of systems embedded on the eating surface such as a table. One such system is a smart table cloth (Fig. 9 (c)) equipped with a fine-grained pressure textile matrix and a weight-sensitive tablet that allowed the spotting and recognition of food intake related actions (such as cutting, scooping, stirring, etc.), the identification of the plate/container on which the action is executed, and the tracking of weight changes in the containers. In the study [97], the authors were able to determine how much food was consumed overall, reporting an error ratio of 16.62% (calculated as error root mean square to signal span) while measuring the food amount consumed.

Another research group employed a stationary weighing sensor [98] and a new algorithm that can detect and measure the weight of individual bites consumed during unrestricted eating. They used an instrumented table (Fig. 9 (b)) with 4 trays placed over food scales. The algorithm works by identifying time periods when the scale weight is stable and then analyzing the surrounding weight changes. The series of preceding and succeeding weight changes were compared against patterns of single food bites, food mass bites, and drink bites to determine if a scale interaction is due to a bite or some other activity. The study involved 271 participants and included a total of ~25000 bites with ground truth labels annotated from video. The weight changes in the scale were used to determine bite portions.

Some studies employed weighing sensors in portable systems. One such study [99] involved a portable smart plate consisting of three load cells. The plate had 3 compartments, each with an embedded load cell (Fig. 9 (d)). The authors here stated that a bite was typically characterized by a sharp increase in measured force originating from food scooping and developed a bite detection algorithm using a random forest decision tree classifier. The authors reported an average error (mean ± standard deviation) of 8 ± 8% in portion size weight estimation in lab conditions. They extended their work in [100] by evaluating a newer bite detection algorithm on multiple measurements with varying food types. The extension also included more realistic eating conditions and a larger dataset. On average, the algorithm was able to estimate the total portion size with an error of 29 g, with an average meal size of 318 g.

Another study [101] presented a similar portable device (the Mandometer, Fig. 9 (a)) for automatically processing continuous in-meal weight measurements to detect in-meal eating indicators, such as total food intake. The algorithm was evaluated on a dataset of 113 meals, using an algorithm that automated the extraction of meal-related indicators. The algorithm calculated the Cumulative Food Intake (CFI) curve of a meal based on continuous weight measurements from the Mandometer. The authors report an error of 24 g for total meal weight.

IV. DISCUSSION

This review was intended to provide a systematic evaluation of existing SB-FPSE methods. The review surveyed 67 full-text research articles, identifying 3 specific sensor types namely wearable, portable, stationary sensors and 2 types of sensor-based methodologies namely direct and indirect FPSE.
Also, 5 sensor modalities were primarily used for FPSE (Acoustic, Strain, Imaging, Motion and Weighing sensors).

Among the three sensor types, most studies in this review made use of portable sensors. These studies employed either portable-imaging or portable-weighing sensors. The ubiquity of cameras in portable devices such as mobile phones has been the main motivation for researchers to use imaging extensively. Mandometer [101] and the smart plate [99], [100] are two portable-weighing sensors.

Stationary sensors are part of a fixed system that is not intended to move. There were two types of stationary sensors, namely, stationary-imaging or stationary-weighing sensors. Stationary-imaging sensors were mostly employed in systems that were fixed in a laboratory. Weighing sensors that are embedded in a fixed eating surface such as a table or tablecloth fall under this category.

The third category of sensors is wearable sensors. Wearable-imaging sensors provide an egocentric point of view and are mostly passive devices. Other wearable sensors, namely wearable acoustical, motion, or strain sensors, mainly sense physiological activities such as chewing, swallowing, and hand/wrist movements. These studies that sense physiological activities provide a different perspective considering metrics related to food intake, unlike the ones that compute the FPS by measuring properties of food (volume, mass).

The review identified that imaging sensors were the most popular of the 5 sensor modalities considered, mostly due to the proliferation of portable imaging on the smartphone platform. The other sensor modalities are often overlooked by the researchers, although they may present compelling alternatives either in stand-alone or hybrid configurations.

For example, a key difference in the sensor operation is whether some form of self-report is required for sensor operation. The self-report may include capture of images by the user of the sensor, placement of dimensional reference in the images, staging of the foods in the image, and others. Reliance on self-report frequently leads to underreporting of food intake [115]. Wearable sensors, on the other hand, may be completely passive, only requiring compliance with wear, but no active self-report by the user. Thus, this review attempts to cover as many sensor modalities as possible, attempting to bring the attention of the research community to less frequently used sensors.

Imaging sensors were broadly classified as single-view and multi-view. The portion size estimation from single-view images relies on dimensional references. Two major concerns pertaining to these methods are assessing irregular shaped food and estimating accurate portion sizes when multiple food items are involved. Another limitation is the use of fiducial markers that are placed in the image scene. Although, these types of references aid in the FPSE, it might not be feasible for the user to place markers in every single eating episode, for example, a snack. If the user fails to place the reference in the scene or take the image, then the whole process becomes void. Some researchers have explored the use of other references such as thumbs, fingers, and circular references (plates/cups/bowls) to eliminate these fiducial markers.

Some other popular methods involve geometric models, shape templates and, virtual wireframes/meshes. They make use of pre-defined shape templates or prior knowledge of the shape of the food item. They fail in the case of irregular shaped food items. Multi-view approaches are the more recent developments in the image-based FPSE milieu and provide more insight into the three-dimensional structure of the food items compared to the single-view approaches. There is no need for fitting shape models nor any prior information such as pre-defined food shapes. Dense-point clouds of the food items can be obtained using multiple perspectives and these methods work well with mixed meals. However, some of the methods still require the use of fiducial markers for calibration and may have issues with non-rigid irregular shaped foods. As the number of views increases, the complexity of the system grows and tends to become cumbersome for the participants.

This review also investigated the viability of using sensor-based approaches for FPSE in free-living, where participants are not limited to the constraints of the lab and can consume any foods at any time. RQ3 was formulated to see whether the methods were tested and validated in the field and represent real-world performance, which may be quite different from that observed in a lab [116]. The overall findings in the review suggest a few important aspects for consideration. These include how the studies reported the measure of accuracy, under what conditions the methods evaluated the performance, what food items were considered, and the reference method that was used. It was observed that the included studies estimate mass/volume and were compared with popular reference ground-truth measures such as water displacement, seed displacement, food weights, and 3D scans.

The first consideration is the accuracy of the SB-FPSE methods. Fig. 10 summarizes the methods that have been reviewed in this article. The review identified that the direct methods for FPSE are more accurate than indirect methods. There is a larger scope for improvement under indirect methods. However, the indirect methods are mostly used with
| Article | View | Image features | Method | Ethical markers | Subjective image capture | Manual input | Pre | Cost | Challenges | Ground Truth | Applicability to live imaging setting | Accuracy error in portion size | Test Data |
|---------|------|----------------|--------|----------------|-------------------------|-------------|-----|-----|-----------|-------------|------------------|-------------------|----------|
| [38]    | Single | iPhone 6s enhanced light setup | Geometry Models, Depth Image based | Yes | Yes | No | Address: foods with unusual shapes, High accuracy with geometry data | Single target, minor Food, Target to outercontour | Irregular shaped foods, Highly reflective background foods | Water Displacement | Yes | No | Paper Cup and 9 NARCO Ohio Repliate |
| [39]    | Single | -SA- | Geometry models, fixed or non-specific camera setup | Yes | Yes | Yes | Improved precision, average correct responses, using optimization | Clothing models | Irregular shaped Foods | -SA- | Yes | No | 40 food types |
| [40]    | Single | Multi phone | Calibration based 3D reconstruction using Ethanol markers | Yes | Yes | Sometimes | Multi app reduces the burden on respondents | Paper only discussed the methods | Validation in the field | -SA- | Yes | No | -SA- |
| [41]    | Single | Nokia 5520 classic | Calibration based 3D reconstruction using Ethanol markers - spherical and prismatic approximation model | Yes | Yes | Yes | Improved model and model inflation impacts volume estimation | Approximation not used for internal food items | Irregular shaped food | Water Displacement | Yes | Three to volume: max 5.5% and min 0.2% | Food items (5 spherical and 2 prismatic -achieved) |
| [42]    | Single | -SA- | Circular reference | No | Yes | No | Works well for regular food items | Irregularly shaped foods are a problem, particular target and thickness of foods | Food items | -SA- | No | -SA- | No food items, including red, uncooked, uncooked |
| [43]    | Single | -SA- | Sphere, fork or chopstick as a reference | Yes | Yes | No | Scattered range of energy from food image | Training dataset is difficult to obtain | Transfer learning | Manual annotation | Average energy estimation error rate of 7.52% | Two container types |
| [44]    | Single | Multi phone | Close up tools (Tablet) | No | Yes | No | Makes use of the performance or self-data to utilize scanning view angles | Necessity to have food colony as a reference | Mixed foods | Food weights | Relative error rate of 5.67% for the weight estimation, Food type (size, manipulation, spikiness, glycemia) |
| [45]    | Single | Multi phone | Table as a reference, abnormal reference | No | Yes | No | Does not have an easy and fast execution | Paper only discussed the methods | Food quantity | -SA- | Yes | No | -SA- |
| [46]    | Single | -SA- | Food shape models using plate food reference | No | Yes | No | Makes use of the performance or self-data to utilize scanning view angles | Blisty food items can be detected | Complex plate food shapes, irregular shaped foods | -SA- | Yes | Error of 3.3% (10.4%), 11.5% for the three tested | Three objects, including a world test, four pieces of skin and a piece of cheese |
| [47]    | Single | Apple iPhone (4S) and Canon Powershot S95 | Shapes Templates | Yes | Yes | No | Accounts | Ethereal Marker | Irregular Shaped Foods | Manual Measurement | 1% error for beans and 1% error for solid (food class) | Real Food |
| [48]    | Single | W稾稶ea Model | No | No | Yes | Works well for various irregularly shaped foods with leases variations | Donut, scones, non-convex objects, scones in case of the known into reference | Irregular shaped foods | Irregularly shaped food | Irregularly shaped food | Average error of 5.98% | 17 food items |
| [49]    | Single | -SA- | Virtual wineglass | No | Yes | Yes | Good accuracy using virtual shapes | Approximate food to regular shapes | Bulk foods and mixed foods with multiple food items | -SA- | Yes | No | -SA- | Tightness regularly shaped objects and food replens |
| [50]    | Single | -SA- | Virtual model for food reconstruction using the Real-World Food App | No | Yes | No | Tabletop-free food preparation, Automatic foods (average absolute error 47.90% (25%)) | Bulk foods and mixed foods | Water displacement or profile measurement | Yes | For large volume foods (average absolute error 16.47% (20%)) | Fifteen food replica |
| [51]    | Single | Phone with depth sensors | EnriC Tang - depth sensor (project stage) | No | Yes | No | Does not use a 3D mark-up tools with the food | All results not reported in the paper | Mixed foods with multiple food items | -SA- | Yes | No | Rice (Duck food) |
| [52]    | Single | RDGLD - handheld handscanner CMF-FPD | Depth sensing | No | Yes | No | Works well for virtual foods, high density 3D into | Build model online into the final stage - integration into the database for 30 foods allowed by 3D blending | Food sparks | Yes | Weight of each food elements within 1% error rate | Food food |
| [53]    | Single | -SA- | Structural Light based 3D reconstruction | No | Yes | No | Works well for the food items in profile | Plastic mode is difficult to real foods | Food foods | Water displacement | No | Due percentage of error for the items tested is around 1% | Two plastic objects including plastic plate |
| [54]    | Single | -SA- | iPhone 4s for the forearm scanner | Automatic 3D reconstruction using a commercial off the shelf TD sensor | No | Yes | No | Reduced sensor burden | Core of the structure sensor | Mixed foods with multiple food items | -SA- | Maybe | In a control setting achieved an average accuracy of 57% | Mixed food items (apple, pear, onion, and pepper and styles of plants were used to validate |
| Device | Function | Description | Portability | Larger | Texture | Calibration | Imaging | References | Notes |
|-------|----------|-------------|-------------|--------|---------|------------|--------|------------|-------|
| [87]  | Microplate | Stained light-based in vivo characterization | No Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [88]  | Depth camera | View consistency approach based on single imaging | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [89]  | Short range depth camera | Depth mapping using water reflection | No Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [90]  | Mobile phone | Stereo-microscopy based on stereo matching | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [91]  | Mobile phone | Finger reference-based volume estimation | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [92]  | Mobile phone | Calibration-based on reference object | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [93]  | iPhone | Shape from silhouettes | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
| [94]  | Google maps | Stereo-microscopy based on stereo matching | Yes Yes No | Works well for different food types and textures | Yes Yes No | Yes Yes No | Opaque | Water displacement | 80% accuracy for in vitro and in vivo analysis |
TABLE II  (Continued.) PORTABLE-IMAGING SENSORS

| Device Type          | Application       | Sensitivity | Specificity | Calibration Method | Validation Method | Paper Quality | Using 3D scans | Advantages | Disadvantages |
|----------------------|-------------------|-------------|-------------|--------------------|-------------------|--------------|----------------|------------|---------------|
| Wearable sensors     |                   | Yes         | Yes         |                    |                   |              | Yes            |            |
|                      |                   |             |             |                   |                   |              |                |            |
|                      |                   |             |             |                   |                   |              |                |            |
|                      |                   |             |             |                   |                   |              |                |            |
|                      |                   |             |             |                   |                   |              |                |            |

wearable sensors and thus, are more friendly to the studies conducted in free-living. It should also be noted that in many studies the accuracy was reported on a few food items, which do not reflect the real-life variability of foods. This creates a paradox: imaging sensors report the highest accuracy rates, followed by weighing sensors with second-best accuracy rates, even though the weighing sensors represent the most direct and standard way of portion size measurement. The difference is easily explained by the conditions (lab vs. free-living) and the size of the studies, both in terms of the number of participants and the number of food items.

The second consideration is the applicability of the reviewed sensors to free-living conditions. The review identified that wearable sensors that indirectly measure portion size (Acoustic, Strain, and Motion) were tested in the laboratory. Further studies are needed to investigate the use of wearable sensors in free-living. The majority of free-living studies to date were based on portable sensors, mostly due to the significant prevalence of smartphones, although some studies used portable-weighing sensors (e.g. [101]).

The review identified several articles that employed hybrid solutions involving multiple sensors. The authors in [31] emphasized the need for devices with multiple sensing modalities. The authors claimed that a combination of modalities led to significant improvements in accuracy. The study proposed a hybrid solution for FPSE using motion and acoustic sensors. An acoustic sensor was used to classify the type of foods and a motion sensor was used to track the number of bites. The authors also hypothesized that acoustic sensors could distinguish between different food textures, for example crisp...
TABLE III

| Article | View | Image Source | Method | Physiological markers | Subjective Image Captures | Manual Input | Pros | Cons | Challenges | Ground Truth | Applicability to free-living setting | Accuracy/ Error in portion size | Test Data |
|---------|------|-------------|--------|-----------------------|----------------------------|--------------|------|-----|------------|--------------|-----------------------------------|-------------------------------|-----------|
| [59]    | Single | Light field camera | Calculating volume using resolution and parameters of the camera | No | Yes | No | Works well for the type of food measured | Densely packed food items are a problem | Different food items | -NA- | No | Error (15.3%) of volume estimation | Apples |
| [60]    | Single | -NA- | Cubic spline interpolation | No | Yes | No | Works well for convex geometries | Requires fine food size in dry portion | Irregular and bulky food items | Shoe method and feature score method | No | Average error in volume 8.425% | Six synthetic asymmetric objects |
| [61]    | Single | Aurora 5 robots: Orions | Depth sensing | No | Yes | No | Works well for irregular food items | Difficult to carry around the robot | Food items | -NA- | No | Total error from the true volume was approximately 5% | White clay-like rice was used as pseudo food |
| [62]    | Two | CO2 Camera | Orthographic projection method | No | Yes | No | Works well for larger volume | Short range, not portable | High pressure: shaped objects (e.g., cubes) and objects that are not axially symmetrical | Water displacement | No | Root mean squared error of volume is 2.586% and 4.625% of the range is shown at arm and high concave-convex objects | 3D images |
| [63]    | Two | RGB-camera | Volume of the food is computed by estimating horizontal and vertical distance of colored image | No | Yes | No | Works well for volumes | Short range, not portable | Different food items | Ellipsoid estimation | No | Estimation error was under 10% for each image | Tomatoes |

*Note: NA = Not Available

TABLE IV

| Article | View | Image Source | Method | Physiological markers | Subjective Image Captures | Manual Input | Pros | Cons | Challenges | Ground Truth | Applicability to free-living setting | Accuracy/ Error in portion size | Test Data |
|---------|------|-------------|--------|-----------------------|----------------------------|--------------|------|-----|------------|--------------|-----------------------------------|-------------------------------|-----------|
| [64]    | Single | View | Infrared camera | Infrared camera and spot light projection | No | No | No | Measurement error was independent of food size | Many foods are not regularly shaped and have irregularities in food size and type | Irregular shaped food and lapse time | Maximum (irregular shaped food) and least displacement | Yes | Average error was 12.05% for the plane method and 20.13% for the LED method. | Few regularly shaped objects and not food-related |
| [65]    | Multi | A robot with a camera | Stereo vision using SLAM | No | Yes | No | No prior information such as pre-defined food size would need | Dense point cloud estimation cannot be done | Unseen surfaces, irregular shaped food | -NA- | Maybe | Overall accuracy of 83% | Snacks, rice cakes, and balconies |
| [66]    | Single | IR camera | Virtual superposed view from different viewpoints | No | Yes | Yes | Improved accuracy compared with visual recognition, which suffers from different positions of food type and position | Irregular shaped food | Food Displacement | Yes | Error rate of 42% with SD of 20.34% | 3D sample sets of women and Asian foods |
| [67]    | Single | Wearable camera | Manual vision/Warped images where possible | No | No | No | Use in free living, single setting | Food must be seen by wearers | Longer duration of the study | Food Recall | Yes | -NA- | 10% foods |
| [68]    | Two | Wearable camera with the last camera | Contrast virtual objects using food and image camera | No | No | No | Works well for regular shaped foods, image recognition using a wearable device, no food camera | Images with irregular shaped food | Bulky foods | Change in overall volume and food displacement monitored by subjects | Yes | Average MAX of 33.1% and 6.46% for food displacement and volume | 18 types of food objects |
| [69]    | Two | Stereo wearable camera | Stereo-based distance estimation | No | No | No | Estimates volume of irregular shaped food items | Stereo processing routines | Bulky foods and nested foods | -NA- | Yes | Overall volume estimation error is only 2.4% with an overall weight estimation error of 2.8% | A larger banana, a pinecone, a fist, chicken wings, and a piece of broccoli cake |

*Note: NA = Not Available

and tacky, and motion sensors could help discriminate between soft foods, such as an ice cream and a milkshake, based on head or wrist position. A similar hybrid solution using features from an acoustic sensor and a strain sensor was discussed in [29]. However, in this case the estimation was prone to errors due to differences in physical properties of different foods.

V. OPEN PROBLEMS AND FUTURE DIRECTIONS

Overall, the sensor-based methods of FPSE made significant progress over the past few years. However, despite promising accuracy, they are not widely adopted in clinical and research practice. There are still open problems and questions to be addressed.

The stationary sensors are limited to the place of installation and thus, are not suitable as a measurement tool for free-living individuals, where the food may be consumed at various places throughout the day. Portable sensors better lend themselves to the idea of free-living monitoring, but these sensors predominantly rely on self-report (active participation of the user) in the form of taking images or placing items on the scale. The limitations of self-report and resulting underreporting of intake are well documented in research literature (although this was not a part of this review). Wearable sensors offer the best alternative in terms of user acceptance and provision of truly “passive” monitoring that does not rely on self-report, only compliance with sensor wear. However, these sensors do not yet provide sufficient accuracy and more work is needed.
to develop methods of portion size estimation from wearable sensor data.

Another significant finding of the review is the very limited scope of sensors that were tested in the applications of portion size estimation. The majority of the reported methods rely on various forms of imaging, with very few other sensor modalities ever tested. The field is ready for innovation in sensor technology, utilizing novel technologies and sensors. For example, time-of-flight cameras may present a compelling alternative to traditional color cameras relying on dimensional references.

The review revealed that the field of FPSE is dominated by imaging sensors with other sensor modalities being underrepresented. A potential reason for such an imbalance is the proliferation of portable imaging devices in smartphones and other consumer electronics. However, the field of FPSE in general is still immature, with a significant number of open problems and directions to grow.

As an indirect outcome of this review, we noticed that a significantly higher number of research publications target the detection of food intake and recognition of food being eaten. A possible explanation is that in the chain of processing to determine energy and nutrient intake, the portion size measurement is important in the later stages, thus FPSE is not considered as the first research priority. We would like to emphasize the extreme importance of the accuracy of the portion size estimation for accurate measurement of energy and nutrient intake and hope that this review will generate a renewed interest in the field.

Although imaging solutions are prevalent, they still have many open problems, such as frequent use of self-report for image capture, need for manual placement of the dimensional image, need for the staging of the foods in the image, and other issues that may affect the process of image capture. Image processing is even more challenging, with errors in FPSE originating from complex food shapes that often need to

---

**TABLE V**

| Article | Device Description | Sensor Type | Sampling Frequency | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|------------------|-------------|-------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [m]    | smart scale equipped with a fingerprint sensor (by weight) and a weigh-motion sensitive scale | Stationary | 40 Hz | spirometer offers high sensitivity, resolution | Depends on certain actions carried out using various sensors, depends, input | Total consumed mass weight force sensors | Field Study | error margin of ±6.2% (calculated as error mean square to signal power) | 5 individuals consumed 8 meals each, 4 meals of each type, 5 times each |

| Article | Device Description | Sensor Type | Sampling Frequency | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|------------------|-------------|-------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [n]    | Universal Eating Monitor | Stationary | 17 Hz down-converted to 1.5 Hz | contact with the weight or individual to determine changes in mass, drinking motion, etc. | Due to the multiple interactions between the size of the food and the person, the interaction is taken to be a fraction of the mass to minimize, and interactions with non-food items | Video Acquisition | Lab Study | error margin of ±3% (calculated as error mean square to signal power) | 251 participants who completed a total of 9,029 meals |

| Article | Device Description | Sensor Type | Sampling Frequency | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|------------------|-------------|-------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [o]    | portable smart phone consisting of three load cells | Portable | 6 Hz | location and magnitude of the number of calories without external sensors | Data captured in a controlled environment, limitations on accuracy | Video Acquisition | Lab Study | error margin of ±3% (calculated as error mean square to signal power) | Two meals |

| Article | Device Description | Sensor Type | Sampling Frequency | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|------------------|-------------|-------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [p]    | Macroseac | Portable | 1 Hz | Force, torque, angle sensors, and sensory inputs | Included variety of foods | CFI curve calculated using sensory/behavioral analysis software | Field Study | error margin of ±3% for total meal weight | 135 meals containing 5,353 kilojoule |

---

**TABLE VI**

| Article | Method | Sensing Modality | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|--------|------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [q]    | Video recording | Acoustic Sensor | Weight prediction models for new foods | Field type, 90 calories were inserted from the recording instrumentation, and was analyzed for correlations between viewing behavior and body weight | Predictive errors in live weights from 10-16% to 19% | Video Acquisition | Lab Study | predictive errors in live weights from 10-16% to 19% | 50 participants and 80% satisfaction of data |

| Article | Method | Sensing Modality | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|--------|------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [r]    | Video recording | Acoustic Sensor | CFI models | Predictive errors in live weights from 10-16% to 19% | Video Acquisition | Lab Study | mean absolute percentage error of 32.2% ± 28.4% for estimating portion size | 30 participants, 44% meals each |

| Article | Method | Sensing Modality | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|--------|------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [s]    | Video recording | Motion and Acoustic Sensor | Behavior analysis models | Did not distinguish between types of foods, or the same types of foods | Video Acquisition | Field Study | error margin of ±3% (calculated as error mean square to signal power) | 144 participants consumed at least two meals of different types and were also housed in a controlled environment |

| Article | Method | Sensing Modality | From | Case | Ground Truth | Type of Study | Accuracy/ Error in portion size | Test Data |
|---------|--------|------------------|------|------|-------------|--------------|-----------------------------|-----------|
| [t]    | Video recording | Smart Sensor | Body composition analysis | Waist circumference measurements, and body fat percentage were used | Video Acquisition, record of tapes of each test was recorded by a scale before and after each test | Lab Study | error margin of ±3% (calculated as error mean square to signal power) | 30 participants each consumed 4 different meals |
be reconstructed from a 2D projection to a 3D representation, food occlusions on plates where one food could completely cover another, and inability to separate ingredients in the mixed dishes, such as soups.

The most significant open problem is in the applicability of any given sensor-based solution to everyday use. We would argue that no existing solution lends itself to everyday use. Weighing and imaging sensors require a significant user burden and may lead to underreporting of the intake. Wearable sensors may require fewer user actions, just cooperation with the wear regiment, but these sensors need to address the issues of accuracy, social acceptance, and data privacy before being widely adopted.

Hybrid solutions that fuse information from several sensor sources may be a promising future direction. However, sensor fusion should not come at the cost of users having to wear multiple sensors resulting in increased sensor burden. Rather, a multi-sensor sensor system may be packaged in socially acceptable accessories such as wrist watches [95], eyeglasses [118], and other wear items. Overall, the future goal should be to achieve an accurate, objective portion size measurement applicable to the wide variety of foods and beverages [118], and other wear items. Overall, the future goal should be to achieve an accurate, objective portion size measurement applicable to the wide variety of foods and beverages consumed by people of different demographic and sociocultural characteristics under the great variety of conditions, environments, and ways to consume food.

VI. CONCLUSION

A comprehensive review of the state-of-the-art SB-FPSE methods was conducted and 5 sensor modalities (Acoustic, Strain, Imaging, Weighing, and Motion sensors) were identified. This article contributed a taxonomy for SB-FPSE methods. The review found that the present-day research is now focusing on improving accuracy, testing outside of restricted laboratory conditions, including mixed meals with more challenging models such as irregular shaped food and non-rigid food items. If these existing challenges can be addressed, SB-FPSE can be exploited to be used in free-living with minimal human intervention in the estimation process.

Indirect methods using wearable technologies can be robust to food shape and size since they are derived from the physiological indicators such as chewing, swallowing, hand gestures, or head movements. The accuracy of FPSE in these methods is lower than with the direct methods. If indirect methods are more extensively explored and the accuracy is improved, they can well be the future of SB-FPSE.

REFERENCES

[1] D. A. Schoeller and M. S. Westerterp-Plantenga, Advances in the Assessment of Dietary Intake. Boca Raton, FL, USA: CRC Press, 2017.

[2] M.-. Labonté et al., “Dietary assessment is a critical element of health research–perspective from the partnership for advancing nutritional and dietary assessment in Canada.” Appl. Physiol., Nutrition, Metabolism, vol. 41, no. 10, pp. 1096–1099, Oct. 2016, doi: 10.1159/apmn-2016-0146.

[3] W. H. Organization. (2020). Healthy Diet. Accessed: Jun. 26, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/healthy-diet.

[4] A. Gulland. (Jan. 2019). Feed the World: Experts Call for Vegetable-Rich Diet to Boost Health and End Hunger. The Telegraph. [Online]. Available: https://www.telegraph.co.uk/news/0/feed-world-experts-call-vegetable-rich-diet-boost-health-end/.

[5] T. V. E. Kral and B. J. Rolls, “Energy density and portion size: Their independent and combined effects on energy intake.” Physiol. Behav., vol. 82, no. 1, pp. 131–138, Aug. 2004, doi: 10.1016/j.physbeh.2004.04.063.

[6] E. Almiron-Roig, J. Solis-Trapala, J. Dodd, and S. A. Jebb, “Estimating food portions. Influence of unit number, meal type and energy density,” Appetite, vol. 71, pp. 95–103, Dec. 2013.

[7] H. T. Kelly et al., “Increased portion size leads to a sustained increase in energy intake over 4 d in normal-weight and overweight men and women,” Brit. J. Nutrition, vol. 102, no. 3, pp. 470–477, Aug. 2009.

[8] K. L. McConahy, H. Smiciklas-Wright, L. L. Birch, D. C. Mitchell, and M. F. Picciano, “Food portions are positively related to energy intake and body weight in early childhood,” J. Pediatrics, vol. 140, no. 3, pp. 340–347, Mar. 2002.

[9] L. K. English et al., “Food portion size and energy density evoke different patterns of brain activation in children.” Amer. J. Clin. Nutrition, vol. 105, no. 2, pp. 295–305, Feb. 2017, doi: 10.3945/ajcn.116.136903.

[10] C. Byrd-Bredbenner and J. Schwartz, “The effect of practical portion size measurement aids on the accuracy of portion size estimates made by young adults.” J. Hum. Nutrition Dietetics, vol. 17, no. 4, pp. 351–357, Aug. 2004, doi: 10.1111/j.1365-277X.2004.00534.x.

[11] S. R. Rapp, P. M. Dubbert, Y. Buttross, and P. A. Burkett, “Reducing diabetics’ food portion size errors by using feedback training,” J. Compliance Health Care, vol. 3, no. 2, pp. 115–124, 1988.

[12] D. L. Sloan and L. H. Eck, “Intense practice enhances accuracy of portion size estimation of amorphous foods,” J. Acad. Nutrition Dietetics, vol. 97, no. 3, p. 295, 1997.

[13] M. Spence et al., “A qualitative study of psychological, social and behavioral barriers to appropriate food portion size control,” Int. J. Behav. Nutrition Phys. Activity, vol. 10, no. 1, p. 92, 2013.

[14] F. Faggiano et al., “Validation of a method for the estimation of food portion size,” Epidemiology, vol. 3, no. 4, pp. 379–382, Jul. 1992, doi: 10.1097/00001648-199207000-00015.

[15] T. Hernández et al., “Portion size estimation and expectation of accuracy.” J. Food Composition Anal., vol. 19, pp. S14–S21, Aug. 2006.

[16] M. Nelson, M. Atkinson, and S. Darbyshire, “Food photography I: The perception of food portion size from photographs,” British J. Nutrition, vol. 72, no. 5, pp. 649–663, Nov. 1994, doi: 10.1079/BJN19940060.

[17] C. S. Venter, U. E. Macht BYU, and H. H. Vorster, “The development and testing of a food portion photograph book for use in an african population,” J. Hum. Nutrition Dietetics, vol. 13, no. 3, pp. 205–218, Jun. 2000.

[18] N. P. Steyn, M. Senekal, S. A. Norris, L. Whati, and J. H. Nel, “How well do adolescents determine portion sizes of foods and beverages?” Asia Pacific J. Clin. Nutrition, vol. 15, no. 1, p. 35, 2006.

[19] N. Alshurafa et al., “Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor.” IEEE Sensors J., vol. 15, no. 7, pp. 3909–3916, Jul. 2015, doi: 10.1109/JSEN.2015.2402652.

[20] J. Cheng, O. Amft, G. Bahle, and P. Lukowicz, “Designing sensitive wearable capacitive sensors for activity recognition,” IEEE Sensors J., vol. 13, no. 10, pp. 3935–3947, Oct. 2013, doi: 10.1109/JSEN.2013.2259693.

[21] M. Farooq and E. Sazonov, “Accelerometer-based detection of food intake in free-living individuals.” IEEE Sensors J., vol. 18, no. 9, pp. 3752–3758, May 2018.

[22] J. M. Fontana, M. Farooq, and E. Sazonov, “Automatic ingestion monitor: A novel wearable device for monitoring of digestive behavior,” IEEE Trans. Biomed. Eng., vol. 61, no. 6, pp. 1772–1779, Jun. 2014, doi: 10.1109/TBME.2014.230673.

[23] T. Vu, F. Lin, N. Alshurafa, and W. Xu, “Wearable food monitoring technologies: A comprehensive review.” Computers, vol. 6, no. 1, p. 4, Jan. 2017, doi: 10.3390/computers6010004.

[24] A. Doulah, M. A. McCrorry, J. A. Higgins, and E. Sazonov, “A systematic review of technology-driven methodologies for estimation of energy intake,” IEEE Access, vol. 7, pp. 49653–49668, 2019, doi: 10.1109/ACCESS.2019.2910308.

[25] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement,” Int. J. Surg., vol. 8, no. 5, pp. 336–341, 2010, doi: 10.1016/j.ijsu.2010.02.007.

[26] F. Fang, S. Aabith, S. Homer-Vanniasinkam, and M. K. Tiwari, “9-High-resolution 3D printed wearable healthcare underrunpin by small-scale fluidics,” in 3D Printing in Medicine, D. M. Kalaskar, Ed. Cambridge, U.K.: Woodhead Publishing, 2017, pp. 167–206.
[106] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, “Measuring calorie and nutrition from food image,” IEEE Trans. Instrum. Meas., vol. 63, no. 8, pp. 1947–1956, Aug. 2014, doi: 10.1109/TIM.2014.2303533.

[107] T. Stutz, R. Dinic, M. Domhardt, and S. Ginzinger, “Can mobile augmented reality systems assist in portion estimation? A user study,” in Proc. IEEE Int. Symp. Mixed Augmented Reality-Media, Art, Social Sci., Humanities Design (ISMAR-MASH’14), Sep. 2014, pp. 51–57, doi: 10.1109/ISMAR-AMH.2014.6935438.

[108] P. Kuhad, A. Yassine, and S. Shimohammadi, “Using distance estimation and deep learning to simplify calibration in food calorie measurement,” in Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA), Jun. 2015, pp. 1–6, doi: 10.1109/CIVEMSA.2015.7158594.

[109] J. Gao, W. Tan, L. Ma, Y. Wang, and W. Tang, “MUSEFood: Multi-sensor-based food volume estimation on smartphones,” in Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (Smart-World/SCALCOM/UIUC/ATC/CRDCom/IOP/SCI), Aug. 2019, pp. 899–906, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00182.

[110] N. Sadeq, F. R. Rahat, A. Rahman, S. I. Ahamed, and M. K. Hasan, “Smartphone-based Calorie Estimation From Food Image Using Distance Information,” in Proc. 5th Int. Conf. Netw., Syst. Secur. (NSyS), Dec. 2019, pp. 1–8, doi: 10.1109/NSyS.2018.863368.

[111] Y. Liu et al., “Food volume estimation based on reference,” in Proc. 4th Int. Conf. Innov. Artif. Intell., New York, NY, USA, 2020, pp. 84–89, doi: 10.1145/3390557.3394123.

[112] F. P.-W. Lo, Y. Sun, J. Qiu, and B. P. L. Lo, “Point2Volume: A vision-based dietary assessment approach using view synthesis,” IEEE Trans. Ind. Informat., vol. 16, no. 1, pp. 577–586, Jan. 2020, doi: 10.1109/TII.2019.2942831.

[113] T. Charoenpong, K. Chamnongthai, P. Kamhom, and M. Krairiksh, “Volume measurement of mango by using 2D ellipse model,” in Proc. IEEE Int. Conf. Ind. Technol., vol. 3, Dec. 2004, pp. 1438–1441. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-27944486523&partnerID=40&md5=efdd94887254d061e2ce1e3de756dfde

[114] S. Ului¸sik, F. Yildiz, and A. T. Özdemir, “Image processing based machine vision system for tomato volume estimation,” in Proc. Electr. Electron., Comput. Sci., Biomed. Eng. ’ Meeting (EBBT), Apr. 2018, pp. 1–4, doi: 10.1109/EBBT.2018.8391460.

[115] M. B. E. Livingstone and A. E. Black, “Markers of the validity of reported energy intake,” Nutrition, vol. 133, no. 3, 2003, pp. 895–920.

[116] A. Doulah, X. Yang, J. Parton, J. A. Higgins, M. A. McCrory, and E. Sazonov, “The importance of field experiments in testing of sensors for dietary assessment and eating behavior monitoring,” in Proc. 40th Annua. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2018, pp. 5759–5762, doi: 10.1109/EMBC.2018.8513623.

[117] A. Doulah, T. Ghosh, D. Hossain, M. H. Imtiaz, and E. Sazonov, “‘Automatic ingestion monitor version 2’—A novel wearable device for automatic food intake detection and passive capture of food images,” IEEE J. Biomed. Health Informat., early access, May 18, 2020, doi: 10.1109/JBHI.2020.2995473.

Viprav B. Raju (Member, IEEE) received the bachelor’s degree in electrical and computer engineering from Visvesvaraya Technological University (VTU), India, and the M.S. degree in electrical engineering from The University of Alabama, Tuscaloosa, AL, USA, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include computer vision processing, sensor networks, machine learning, and deep learning. His current research interests include dietary assessment and image-based food intake monitoring.

Edward Sazonov (Senior Member, IEEE) received the Diploma degree in systems engineering from the Khabarovsk State University of Technology, Russia, in 1993, and the Ph.D. degree in computer engineering from West Virginia University, Morgantown, WV, USA, in 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA, and the Head of the Computer Laboratory of Ambient and Wearable Systems. His research interests include wireless, ambient, and wearable devices, and methods of biomedical signal processing, and pattern recognition. Devices developed in his laboratory include: a wearable sensor for objective detection and characterization of food intake; a highly accurate physical activity and gait monitor integrated into a shoe insole; a wearable sensor system for monitoring of cigarette smoking; and others. His research has been supported by the National Science Foundation, the National Institutes of Health, the National Academies of Science, as well as by state agencies and private industry and foundations.