

# Food Volume Estimation for Quantifying Dietary Intake with a Wearable Camera

Anqi Gao, Frank P.-W. Lo and Benny Lo

**Abstract**—A novel food volume measurement technique is proposed in this paper for accurate quantification of the daily dietary intake of the user. The technique is based on simultaneous localisation and mapping (SLAM), a modified version of convex hull algorithm, and a 3D mesh object reconstruction technique. This paper explores the feasibility of applying SLAM techniques for continuous food volume measurement with a monocular wearable camera. A sparse map will be generated by SLAM after capturing the images of the food item with the camera and the multiple convex hull algorithm is applied to form a 3D mesh object. The volume of the target object can then be computed based on the mesh object. Compared to previous volume measurement techniques, the proposed method can measure the food volume continuously with no prior information such as pre-defined food shape model. Experiments have been carried out to evaluate this new technique and showed the feasibility and accuracy of the proposed algorithm in measuring food volume.

## I. INTRODUCTION

Previous health surveys in England reported that 65% of men and 58% of women were overweight in 2014 [1]. Unhealthy dietary intake, including unbalanced dietary pattern and excess calorie intake, is one of the major factors which leads to obesity [2]. A daily dietary assessment system can potentially help users and dieticians to understand their dietary behaviour with information about calorie and nutrient intake. The traditional method of dietary assessment which has long been relied upon is user self-report, such as 24 hour dietary recall. Users need to report their food intake with detailed information about consumed weight or volume. Such subjective measurement are highly inaccurate, and user compliance is also a major issue. Moreover, the consumed weight or volume reported highly depends on users' subjective judgement which may also lead to a biased dietary analysis result. In recent years, wearable devices have become popular for long-term health monitoring [3], [4]. Increasing numbers of people have become health conscious and started using their smart-watches or fitness bands for personal daily exercise recording and analysis. The low cost and pervasiveness of the wearable technologies have made them attractive for healthcare applications. Hence, a wearable dietary monitoring device could be developed to address the need for objective dietary analysis.

A complete procedure of quantifying dietary intake consists of food detection and segmentation, consumed

weight/volume estimation, nutrient intake calculations, and dietary analysis. For food detection and segmentation part, it can be performed using convolution neural network (CNN) [5]. For example, Google applied its generic food detection with the GoogLeNet model (a deep neural network) and achieved the mean average precision of 80% [6]. For volume estimation, previous research studies mostly rely on model-based techniques or 3D model generation based on back projection using camera calibration matrix [7]. To our best knowledge, there is no published work on using a monocular SLAM system for food volume measurement. This project explores a novel way of using SLAM to estimate the consumed food volume which could be embedded easily into a wearable camera device. Once the food volume is measured by the wearable device, the data can be merged with USDA national nutrient database for further dietary analysis [5].

## II. BASIC THEORY

### A. The Visual SLAM Framework

The visual SLAM framework can be divided into four parts including visual odometry, loop closure, back-end optimisation and mapping [8]. Visual odometry aims to estimate the camera's pose by analyzing the feature points on the captured image between different views, so that the trajectory of the moving camera can be computed. Loop closure is the process of recognizing the location which has been previously captured in order to correct the drift trajectory of the camera. Back-end optimisation processes information from visual odometry and loop closure, and performs both local and global optimisation for localisation and mapping based on different algorithms. Once obtained the optimised camera's trajectory, mapping is used to build an environment map (sparse map in this paper) from captured image sequence. The sparse map consists of point clouds which associate with the feature points extracted from the images. Fig.1 shows the sparse map of a can of soft drink generated by the SLAM system using the feature points.

### B. 3D Mesh Object Reconstruction

3D mesh object reconstruction is used to construct a 3D mesh from the point clouds which represent the shape of the target object. There are several 3D mesh reconstruction algorithms proposed including greedy triangulation, grid projection and Poisson [9]. Though some of them do have great performance in volume measurement after reconstruction, these techniques involve parameters which require manual intervention to get a well-reconstructed mesh. For example, Poisson surface reconstruction shows a high accuracy in

A. Q. Gao, P. W. Lo and B. Lo are with the the Hamlyn Centre, Imperial College London, UK, e-mail:{anqi.gao16, po.lo15, benny.lo}@imperial.ac.uk. This project is partially supported by Lee Family Scholarship, awarded to P. W. Lo, and B&M Gates Foundation funded project "An Innovative Passive Dietary Monitoring System" (OPP1171395).

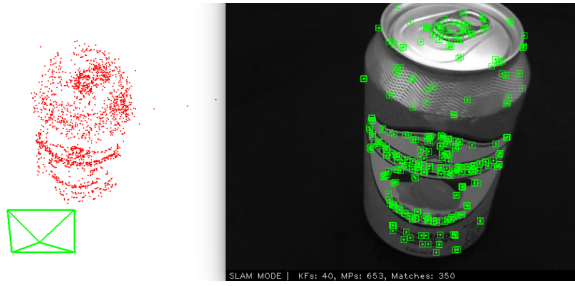


Fig. 1: The left image shows the generated sparse map; The right one shows an image captured by the camera with feature points.

further volume measurement, however, it is not adaptive to an automatic SLAM system. In order to develop a technique which can be integrated with the SLAM system, convex hull algorithm, one of the 3D reconstruction methods, seems to be the best option among those previously mentioned techniques. Moreover, in order to improve the accuracy of food volume measurement, a further optimisation on convex hull algorithm is developed. The detailed information will be presented in the following sections.

### III. DETAILED INFORMATION AND METHODS

#### A. Statistical Outlier Removal Filter

After the sparse map is generated by the proposed SLAM system, there are always inevitable outliers. In monocular SLAM, the depth information is not known and this has to be estimated from the images. An incorrect depth estimation or environment noise will induce outliers to the sparse map as shown in Fig.2a. Moreover, outliers will lead to overestimation in volume measurement. The detailed method of statistical outlier removal filter is shown in Algorithm 1. With the use of the proposed method, most of the outliers will be trimmed from the point cloud as shown in Fig. 2b.

---

#### Algorithm 1: Statistical Outlier Removal Filter

---

```

1 // point_distance[i] is the average distance to k nearest
  neighbours for each point ;
2 // Initialize result[i] to be TRUE ;
3 // n is a multiplication factor of the standard deviation;
4 meanStdDev(point_distance,mean,standard_variance);
5 for (i = 0; i < points.size(); i++) do
6   if result[i] then
7     result[i] = point_distance[i] <
      mean + n * standard_variance;
8   end
9 end
10 return result

```

---

#### B. Point Cloud Completion

The point cloud generated by the SLAM system often loses information due to a limited viewing angle during image capturing, and this will lead to underestimation of the food volume. Since there is no published materials which use

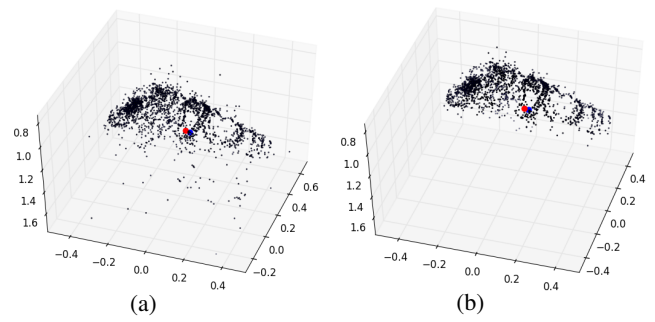


Fig. 2: Sparse map generated by SLAM system: (a) Point cloud of a can with outliers; (b) Point cloud of a can without outliers

SLAM for volume estimation, our work will firstly be carried out based on several assumptions. First, we assume that the detected food is solid and central symmetric so that its shape can be completed based on symmetry. Second, the top, front and side surfaces (3 out of 6 if a cube is considered) should be acquired during video capturing. With those assumptions, 3D object reconstruction can be carried out. Assume the original point cloud is  $P$ , there are  $n$  points in the point cloud  $P$ . The centroid  $p_m$  of the point cloud can be calculated:  $p_m = \frac{1}{n} \sum_{i=1}^n p_i$  for  $i = 1, 2, \dots, n$ . The detailed procedure is shown in Algorithm 2. After point cloud completion, the filtered point cloud can be completed as shown in Fig.3.

---

#### Algorithm 2: Point Cloud Completion

---

```

1 // points contain the points in point cloud P;
2 // new_points is the point cloud  $P_{new}$  with new points
  added;
3  $p_m = \text{find\_centroid}(\text{points})$  ;
4 // Completion based on symmetry;
5 for (point IN points) do
6   | new_points.add( $2 * p_m - \text{point}$ ) ;
7 end
8 return new_points

```

---

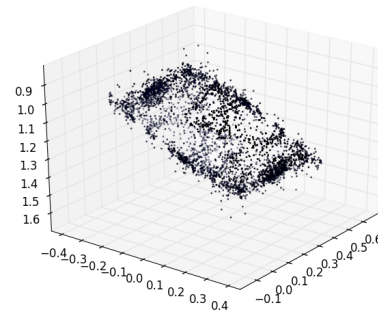


Fig. 3: The 3D structure of a soft drink can generated by the SLAM system with point cloud completion

#### C. Convex Hull with Boundary Shrinking Property

Convex hull algorithm is a well known 3D mesh reconstruction method and which has been applied in our technique. Fig. 4a and Fig.4b show the 3D mesh reconstruction of 2 objects by 3D Quick Hull [10]. In order to generate

the 3D mesh accurately with convex hull, the target object should be convex, otherwise, the algorithm will fill up all the non-convex space which leads to overestimation of the object volume. Hence, another assumption in our technique is that the target object has to be convex. In mathematics, the convex hull  $C$  of a point set  $P$  in the Euclidean space is the smallest convex sets which is able to enclose all the points in  $P$  [11]. However, convex hull algorithm often overestimates the volume of the reconstructed mesh in 3D reconstruction leads to a larger mesh compared to the original target object. The reason is that convex hull is sensitive to noise as well as outliers. This will cause inaccuracy in reconstructing the shape of the object. Hence, a novel boundary shrinking method has been introduced into the proposed technique. The shrinking method is an exhaustive search method where convex hull is applied multiple times to find the optimal object shape with the larger numbers of vertices (boundary points) encapsulated. The detailed method is shown in Algorithm 3. After the boundary shrinking, the volume of the point cloud  $P_{max}$  can be computed by using the volume measurement function in PCL library [10].

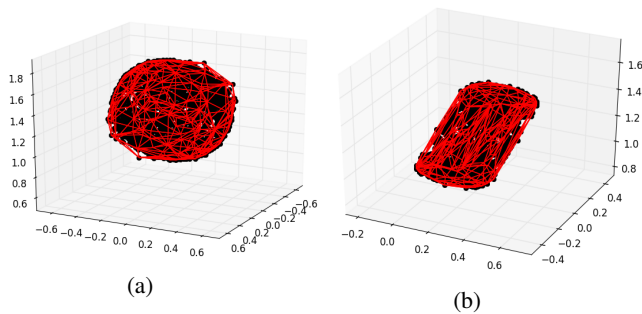


Fig. 4: 3D mesh reconstruction of 2 objects using 3D Quick Hull: (a) An apple reconstructed from its point cloud (b) A soft drink can reconstructed from its point cloud

---

### Algorithm 3: Boundary Shrinking Method

---

```

1 //  $P_0$  contains all points in the original point cloud;
2 //  $V_0$  contains the vertices of the original point cloud;
3 //  $find\_vertices()$  is the function to find the vertices of
  a set of points;
4 //  $max$  is to temporarily store the iteration with the
  largest number of vertices;
5 for ( $i = 0$ ;  $i < num$ ;  $i++$ ) do
6    $P_{i+1} = P_i - V_i$ ;
7   if  $V_i.size() > V_{max}.size()$  then
8      $max = i$ ;
9   end
10   $V_{i+1} = find\_vertices(P_{i+1})$ ;
11 end
12 return  $P_{max}$ ;

```

---

## IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

To evaluate the performance of the proposed technique, several experiments have been carried out to assess if the

proposed SLAM system combined with convex hull algorithm can accurately estimate food volume. To explore the possibility of continuous food volume measurement with a monocular camera, an Apple iPhone 6 plus and a 4k wearable action camera have both been used for data collection. The frame rate and the dimension of the video have been set to 30fps and 640x480 pixels. In the process of capturing video data, target object has been placed on a totally black background. A Rubiks cube has been designed as a scale reference for calibration.

### A. The Performance of Boundary Shrinking Property

Several target objects have been used to evaluate the performance of proposed boundary shrinking method. The results of two (an apple and a soft drink can) have been presented in the following. The exact volume for them are  $230\text{ cm}^3$  and  $350\text{ cm}^3$  respectively. Fig.5 and Fig.6 present the volume of the target objects estimated after boundary shrinking. In the proposed method, the mesh with the largest number of vertices will be selected for volume measurement. The experiment results show that the meshes constructed with 6 iterations have the largest number of vertices for both objects. The percentage error for the apple and the can is 10.03% and 10.89% respectively. It has a much higher accuracy compared to the one using the traditional convex hull algorithm (the percentage error is 31.53% and 28.49% for the apple and the can respectively). After the experiment, it is reasonable to conclude that the proposed method is robust to outliers and demonstrated that it is feasible to use SLAM for continuous volume measurement.

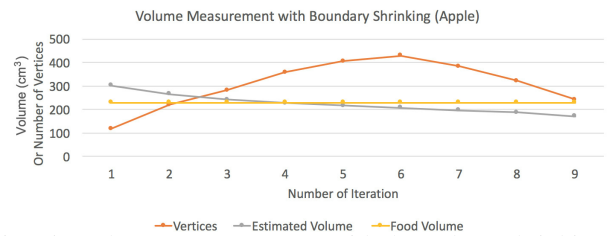


Fig. 5: Volume measurement with boundary shrinking in different iterations (apple)

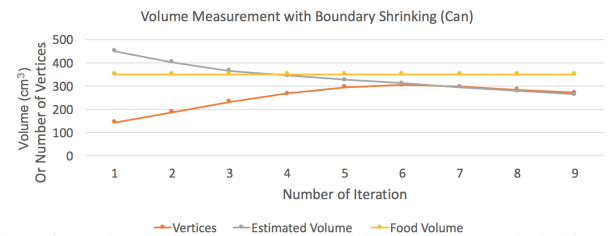


Fig. 6: Volume measurement with boundary shrinking in different iterations (can)

### B. Food Volume Measurement on Static Food

In this experiment, a mini-cake with the volume  $60\text{ cm}^3$  has been recorded. First, the measurement reference, Rubiks cube, has been recorded for initialising the SLAM system. Afterwards, the camera is moved to the cake for scanning, and a camera trajectory is shown in Figure 7. The volume of the target object for each moment has been computed with

the use of captured images from the corresponding timestamps (23 timestamps in the experiment). The final volume is measured by taking average of the volume recorded over the timestamps. The experiments have been repeated several times for reliability. The percentage errors of the volume measurement over the experiments for static mini-cake are  $7.78 \pm 1.7\%$ ,  $19.67 \pm 1.9\%$  and  $18.20 \pm 3\%$  respectively. The average percentage error is shown as  $15.21\%$ .

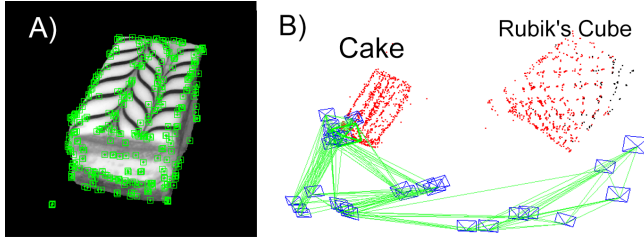


Fig. 7: Food volume measurement on static food: A) Tracking the cake; B) Generated sparse map; The green line is the estimated camera trajectory by the SLAM system

The first experiment shows the best performance in volume measurement among three trials. The second and third experiments show a relatively high percentage error compared to the first one. From the 95% confidence interval among the experiments, it is reasonable to say that the SLAM system is stable in volume measurement. Since the scale for the point cloud is relative in monocular SLAM, this is the reason why a Rubik's Cube is needed for initialisation at the beginning. The reason for the percentage error could be due to the uncertainty in the SLAM system initialisation.

### C. Food Volume Estimation During Food Consumption

In order to explore the feasibility of the proposed technique in continuous measurement of food consumption, a sausage is used in an experiment. The volume of the sausage is  $81 \text{ cm}^3$ . As shown in Fig. 8, the sausage has been cut and small sections have been taken away one by one to simulate the process of eating until there is only one left behind. To ensure the reliability of the experimental results, the experiment has been repeated several times. The result is shown in Fig.9. It can be seen that the estimated volume decrease when the sausage is taken away. The percentage errors of the volume measurement over the experiments are  $23.9 \pm 1.8\%$ ,  $25.40 \pm 2.2\%$  and  $19.39 \pm 2.3\%$  respectively. The average percentage error is shown as  $22.8\%$ . Further details on the results are also shown in Table I.

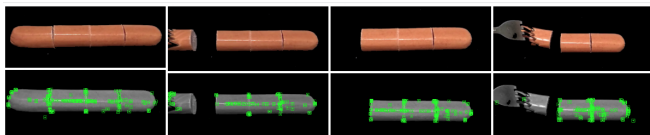


Fig. 8: Frames captured for food volume monitoring during food consumption

## V. CONCLUSION

This paper has introduced a new concept of using a monocular vision based SLAM system to estimate food

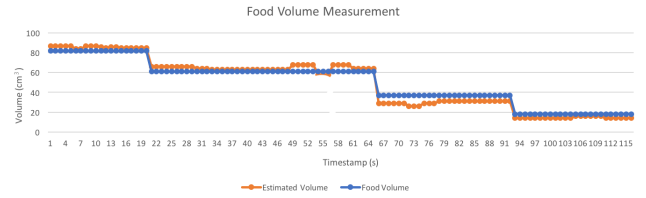


Fig. 9: Experimental result of food volume estimation during the food consumption (i.e. the sausage)

TABLE I: Percentage Error (%) for Volume Measurement

Food	1st experiment	2nd experiment	3rd experiment	Average
Mini-cake (iPhone)	$7.78 \pm 1.7\%$	$19.67 \pm 1.9\%$	$18.20 \pm 3.0\%$	$15.21\%$
Mini-cake (Action Cam)	$13.46 \pm 2.1\%$	$22.67 \pm 2.0\%$	$17.30 \pm 1.9\%$	$17.81\%$
Sandwich (iPhone)	$14.45 \pm 2.0\%$	$6.37 \pm 1.7\%$	$14.25 \pm 2.2\%$	$11.69\%$
Sandwich (Action Cam)	$14.73 \pm 1.7\%$	$23.47 \pm 2.1\%$	$19.50 \pm 3.0\%$	$19.20\%$
Food consumption	1st experiment	2nd experiment	3rd experiment	Average
Sausage (iPhone)	$23.9 \pm 1.8\%$	$25.40 \pm 2.2\%$	$19.39 \pm 2.3\%$	$22.8\%$
Sausage (Action Cam)	$25.7 \pm 2.0\%$	$31.6 \pm 1.5\%$	$26.4 \pm 1.7\%$	$27.90\%$
Mini-cake (iPhone)	$18.31 \pm 1.6\%$	$20.70 \pm 2.5\%$	$14.39 \pm 2.2\%$	$17.80\%$
Mini-cake (Action Cam)	$20.42 \pm 2.4\%$	$19.42 \pm 2.3\%$	$16.33 \pm 2.5\%$	$18.70\%$
Sandwich (iPhone)	$17.19 \pm 1.9\%$	$18.45 \pm 2.4\%$	$13.32 \pm 2.2\%$	$16.32\%$
Sandwich (Action Cam)	$20.33 \pm 1.4\%$	$23.53 \pm 2.5\%$	$15.49 \pm 1.9\%$	$19.70\%$

volume dynamically. The proposed technique shows the feasibility and accuracy in continuous food consumption measurement. With the use of the statistical outlier filter, the point completion technique and the multiple convex hull algorithm, the proposed technique can get a performance with an overall accuracy of  $83\%$ .

## REFERENCES

- [1] Lifestyles Statistics Team and Paul Niblett. Statistics on obesity, physical activity and diet. *Health and Social Care Information Centre, London*, 2016.
- [2] George A Bray and Barry M Popkin. Dietary fat intake does affect obesity! *The American journal of clinical nutrition*, 68(6):1157–1173, 1998.
- [3] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G. Z. Yang. An intelligent food-intake monitoring system using wearable sensors. In *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, pages 154–160, May 2012.
- [4] Claire Pettitt, Jindong Liu, Richard M. Kwasnicki, Guang-Zhong Yang, Thomas Preston, and Gary Frost. A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. *British Journal of Nutrition*, 115(1):160167, 2016.
- [5] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1085–1088. ACM, 2014.
- [6] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. Im2calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1233–1241, 2015.
- [7] Chang Xu, Ye He, Nitin Khannan, Albert Parra, Carol Boushey, and Edward Delp. Image-based food volume estimation. In *Proceedings of the 5th international workshop on Multimedia for cooking & eating activities*, pages 75–80. ACM, 2013.
- [8] Xiang Gao, Tao Zhang, Yi Liu, and Qinrui Yan. *14 Lectures on Visual SLAM: From Theory to Practice*. Publishing House of Electronics Industry, 2017.
- [9] Frederik V. Berentsen, Andrea Keiser, Ann-Marie H. B. Bech. Mesh reconstruction using the point cloud library. 2015.
- [10] pointclouds.org. Documentation - point cloud library (pcl). 2014.
- [11] Franco P. Preparata and Se June Hong. Convex hulls of finite sets of points in two and three dimensions. *Communications of the ACM*, 20(2):87–93, 1977.