**Supplementary Material:**

**S1. System Calibration**

During the system calibration process, we first put the calibration pattern, which is a checkerboard, on top of the turntable. We choose a point on one corner of the checkerboard as the origin of the world coordinate system (Fig.2(a)), and the checkerboard in the same plane of the  plane of the world coordinate system. Then, the axis perpendicular to the turntable is the  axis of the world coordinate system (Fig.2). The standard calibration algorithm (1) (its mathematical description is omitted here) requires two sets of parameters. The first set consists of intrinsic parameters, such as the focal length of the camera lens and the pitch size of the sensor chip. The second set contains extrinsic parameters, including the relative position and pose information between different viewpoints of the cameras.

**S2. Sparse 3D Point Cloud Reconstruction of Multi-View Stereo Images**

We use the multi-view reconstruction method to recover three-dimensional information of food. The well-established algorithms of SURF (2), Harris (3) and FAST (4) with combined features were implemented to extract feature points from images. SIFT(5, 6) vectors are chosen as the feature descriptors of the extracted feature points, and the feature vectors whose Euclidean distance is larger than a preset threshold are regarded as a pair of matched feature points. Once the matched feature point tracks are obtained, we use the triangulation method to calculate the coordinates of the 3D points (7). The feature point track is also known as the feature string, in which the length (i.e., the number of features included in the feature string) is an important indicator of the reliability of the feature track. One advantage of the multi-view reconstruction method is that the three-dimensional scene or target can be observed by multiple perspective views. This advantage is reflected in the sparse point cloud reconstruction, which can obtain a large number of relatively long feature strings, indicating that the feature represented by the feature string can be likely matched in many views. In addition, a longer feature string implies higher feature distinguishing power and saliency, and a higher resistance to the matching error. Due to the preference of longer feature strings, we filter out the shorter ones, i.e., if the feature string of a 3D point  is , **M** will be eliminated when

 (1)

where  is the length of , and  is a threshold.

The homogeneous coordinate of 3D point  in the world coordinate system corresponds to a point  in each of the *l* image coordinate systems (Fig.3). The projection matrix between the two coordinate systems is , where **K**, **R** and **t** are the calibrated camera parameters. The relationship between image coordinates and the corresponding world coordinate  can be represented as a rigid body transformation

 (2)

where  is an arbitrary non-zero scalar determined according to the perspective projection, and  is a 4×3 matrix, given by

 (3)

Substituting (3) into (2), we have

(4)

For simplification, Eq. (4) are re-expressed by defining **A, b** and , given by

 , ,  (5)

Then (4) becomes

 (6)

The least square solution for  is given by

 (7)

Due to unavoidable errors in camera calibration and feature matching, noisy points are present in the point cloud after implementing (7). In order to suppress noise, all points in the cloud are checked using the algorithm described in (8). A point is considered to be an outlier if , where is the average distance to its  nearest neighbors, and is a preset threshold which is chosen to be one standard deviation from the mean of average distance to all neighbor points.

**S3. Volume Estimation Based on Electric-Field Physical Model**

We present two methods for food volume estimation, a simple sliced point cloud method and a robust estimation method using a new electric field based physical model.

**S3.1 Sliced Point Method**

In this method, we first slice the point cloud in equal-thickness layers along a given axis (e.g., the *z*-axis). Then, each layer of points is projected to the plane (e.g., the *x*-*y* plane) perpendicular to the given axis, resulting in sub-cloud point sets, :

 (8)

We calculate the area  enclosed by point set and integrate along the direction of the given axis to estimate volume:

 (9)

whereis the layer thickness, and *V* is the estimated volume of food.

**S3.2**  **Method Based on Electric Field Physical Model**

In order for Eq. (9) to produce accurate volumetric estimation, the boundary of each slide needs to be adjusted or smoothed according to the type of food and the characteristics of the point cloud. For an arbitrary object in 3D space, its surface can be divided into a finite number of triangular patches which form polyhedrons. Although it is desirable to use the summed volumes of spatial polyhedrons, which can be calculated conveniently, as the estimate of the food volume, an assumption has to be made that the adjacencies within the point cloud constituting the spatial polyhedrons are known in advance. Unfortunately, the neighborhood information of these spatial points cannot be obtained easily from the three-dimensional point cloud established via the multi-view reconstruction technique. Thus, it is difficult to apply the spatial polyhedral volume calculation formula to the point cloud. In order to solve this problem, we propose an electric field based physical model that simulates the action of charged particles. When applying this model to our volume estimation problem, the food surface, which is either locally convex or concave, is searched automatically, based on a physical model that charged particles with mobility gradually approach the surface of the domain where the particles are contained due the force of the electric-field applied to charged particles. Since the adjacencies between "charged particles" are known, our method greatly reduces the difficulty in finding the triangular meshes directly from the point cloud. Once the adjacencies are obtained, we can then apply the polyhedral volume calculation formula.

Based on this new concept, we hypothetically assume that the reconstructed food point cloud as a set of Positively Charged Particles (*PCP*). The positions of all particles in the *PCP* set are fixed (stationary). We also hypothetically define another set of Negatively Charged Particles (*NCP*). In contrast to *PCP*, the particles in set *NCP* are free, i.e., these particles are free to move under the electric field force.

The first step is to initialize the particles in *NCP*. Similar to slice-based method(9, 10), we slide the food point cloud into  layers along a given axis and project each layer of points into the plane perpendicular to the given axis. In our case, we choose the axis. As a result,  sub-cloud point sets are obtained by .

For each , we slide it again along the *z*-axis but in a perpendicular direction, dividing  into strips. The centers of all the strips, , are calculated, given by

 (10)

Then, the point at the middle of set , denoted by , is found. Next, we traverse all points in  calculating their Euclidean distances from , and the minimum value is denoted as . We define  as the mean value of the *z* coordinate component in point set and in the meantime we choose  as the center and  as the radius to build a circular point set in the plane of , which is denoted as (Fig.4). We also define the middle point of the point set  as.

In this way, we have obtained the initial positions of the particles in *NCP* with respect to the food point cloud. According to the electric field theory, there exists an electric field produced by the charged particles in *PCP* and *NCP*. For a single particle  and a single particle , the electric field force between them is given by

 (11)

where  is the distance between  and  (Fig.4), *K* is a coefficient which we define as  to make Eq. (11) consistent with the Coulombs law(11). We take the amount of charge in one particle in *NCP* as the unit charge. Therefore the amount of charge carried by one is  (). The amount of charge carried by one  is , which is decided by:

 (12)

where  represents the convex volume formed by food point cloud, and  is the number of 3D points in food point cloud.

According to the kinetic energy theorem(12), a free particle in the *NCP* set will accelerate proportionally to the electric field force, and its kinetic energy is quadratically proportional to the speed of movement. As the negatively charged particle moves closer to the positively charged ones (stationary particles in the *PCP* set), its kinetic energy increases and reaches the maximum in the region where the point cloud is the densest. Therefore, we define the contour of the *j*th strip as  at the points where the kinetic energies of particles in *NCP* are maximized. For simplicity, we assume that each particle in *NCP* moves within a virtual pipe whose direction is represented by a vector , where  is the *j*th point in set. The virtual pipe assumption promotes particle movement in the direction and limits movements in other directions.

The final positions of particles in *NCP* are given by :

 (13)

where is the total number of 3D points in set *NCP*. The food volume is calculated based on *con*. Since the *NCP* set is constructed in advance, the adjacencies of particles are known. Accordingly, the triangular mesh model of the food can be obtained.

The positive determinant for each triangular mesh is calculated by the three points , andof the mesh being arranged in the counterclockwise order (from the outside to the inside) as follows:

 (14)

where, and .Then, the volume enclosed by the entire triangular meshes can be expressed as:

 (15)

where *V* is the measured food volume placed on the VD meter, is the determinant of themesh and  is the total number of triangular meshes.

**References:**

1. Tsai R. Y. (1987) A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. on Robotics and Automation* 3(4), 323-344.
2. Bay H., Tuytelaars T. & Gool L. J. V. (2006) SURF: Speeded Up Robust Features. *Proceedings of the European Conference on Computer Vision*, pp. 404-41.
3. Harris C. & Stephens M. (1988) A combined corner and edge detector. *Proceedings of Alvey Vision Conf*.*,* pp. 189-192.
4. Rosten E. & Drummond T. (2006) Machine learning for highspeed corner detection. *Proceedings of the* *European Conf. on Computer Vision*, pp. 430-443.
5. Lowe D. G. (2004) Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60(2), 91-110.
6. Lowe D. G. (1999) Object recognition from local scale-invariant features. *Proceedings of the Int. Conf. on Comput. Vision*, pp. 1150-1157.
7. Ma Y., Soatto S., Košecká S. et al. (2004) An invitation to 3-D vision: from images to geometric models. Springer Verlag.
8. Rusu R. B. Marton Z. C., Blodow N. et al. (2008) Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems* 56(11), 927-941.
9. Dehais J., Shevchik S., Diem P. et al. (2013) Food volume computation for self-dietary assessment applications. *Proceedings of the* *IEEE Int. Conf. on Bioinformatics and Bioengineering (BIBE)*, pp. 1-4.
10. Dehais J., Anthimopoulos M., Shevchik S. et al. (2017) Two-view 3D reconstruction for food volume estimation. *IEEE Trans. Multimedia* 19(5), 1090-1099.
11. Hetherington D. (1997) Coulomb’s law. *Phys. Educ.* 32(4): 277.
12. Chu L. J. (1951) A Kinetic Power Theorem. *Presented at the IRE‐PGED Electron Tube Research Conference, Durham, New Hampshire, June.*