Food Volume Estimation in a Mobile Phone Based Dietary Assessment System

Md Hafizur Rahman[a], Qiang Li[a], Mark Pickering[a],
Michael Frater[a], Deborah Kerr[b]
[a]School of Engineering & Information Technology
The University of New South Wales
Canberra, Australia
e-mail: md.rahman2@student.adfa.edu.au

[b]School of Public Health
Curtin University
Perth, Australia

Carol Bouchey[c], Edward Delp[d]
[c]University of Hawaii Cancer Center
University of Hawaii
Hawaii, USA

[d]School of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
e-mail: ace@ecn.purdue.edu

Abstract— There is now convincing evidence that poor diet, in combination with physical inactivity are key determinants of an individual's risk of developing chronic diseases, such as obesity, cancer, cardiovascular disease or diabetes. Assessing what people eat is fundamental to establishing the link between diet and disease. Food records are considered the best approach for assessing energy intake. However, this method requires literate and highly motivated subjects and adolescents and young adults are the least likely to undertake food records. The ready access of the majority of the population to mobile phones has opened up new opportunities for dietary assessment. In such systems, the camera in the mobile phone is used for capturing images of food consumed and these images are then processed to automatically estimate the nutritional content of the food. A vital step in this process is the estimation of the volume of the food in the image. In this paper we propose a food volume estimation approach which requires only a pair of stereo images to be captured. Our experimental results show that the proposed approach can provide an accurate estimate of the volume of typical food items in a passive manner without the need for manual fitting of 3D models to the food items.

Keywords- dietary assessment; food records; feature detection; volume estimation; disparity map; depth map; 3D point cloud

I. INTRODUCTION

Preventing disease through improving nutrition is a global health priority [1]. Approximately 30% of all cancers have been attributed to dietary factors [2]. The strongest evidence for diet increasing cancer risk is specifically with overweight and obesity, high consumption of alcoholic beverages, aflatoxins and fermented foods. A diet of at least 400 g per day of fruits and vegetables appears to decrease cancer risk. However, a key barrier to linking dietary exposure and disease is the ability to measure dietary factors, including intake of food groups such as fruits and vegetables, with specificity and precision [3].

Assessing what people eat is fundamental to establishing the link between diet and disease. However, it is now more challenging to do this as consumers have moved away from eating a traditional 'meat and 3-veg' meal at home to purchasing more take-away food and eating out [4, 5]. With this greater proportion of foods eaten away from home [6, 7], it is now becoming increasingly difficult for consumers to accurately assess how much they have eaten or the composition of their meal.

Food records are considered the best approach for assessing energy intake (kilojoules). With a paper-based food record, subjects are asked to record their food and fluid intake for between 3-7 days. This method requires literate and highly motivated subjects. Research has shown adolescents and young adults, who typically have unstructured eating habits and frequently snack, are the least likely to undertake food records [8].

With advances in technology it is now timely to explore how mobile devices can better capture food intake in real-time by potentially reducing the burden of the recording task to both the subject and the researcher. The ready access of the majority of the population to mobile phones has opened up new opportunities for dietary assessment which are yet to be leveraged. Tufano et al. [9] in a review of eHealth (web and mobile phone) applications refers to this as 'technology convergence' in which real-time or near-real time multimedia communication capabilities can occur.

The integrated camera in the mobile phone is used for capturing images of food consumed. These images are then processed to automatically estimate the nutritional content of the food items for record keeping purposes and to provide feedback to the patient. To estimate the nutritional content of food items in an image the food item must be recognized and the mass of the food item must be estimated. A vital step in estimating the mass of the food item is to estimate its volume as this can then be used in conjunction with a density database of food items to estimate the mass of the food in the image. Previous approaches to volume estimation of food have included passive and active approaches which require the user to capture from one image up to several pairs of images of the food item.

Shang et al. [10, 11] proposed an active approach to food volume estimation using structured light. In their proposed system, a laser module is attached to a mobile phone. This module produces a rectangular grid pattern with the brightness of the lines decreasing according to the distance from the center of the pattern. Grid lines are extracted from the camera image and a depth map is created from multiple pairs of images after a calibration stage. This depth map can be used to create a 3D surface

978-0-7695-4911-8/12 $26.00 © 2012 IEEE
DOI 10.1109/SITIS.2012.146

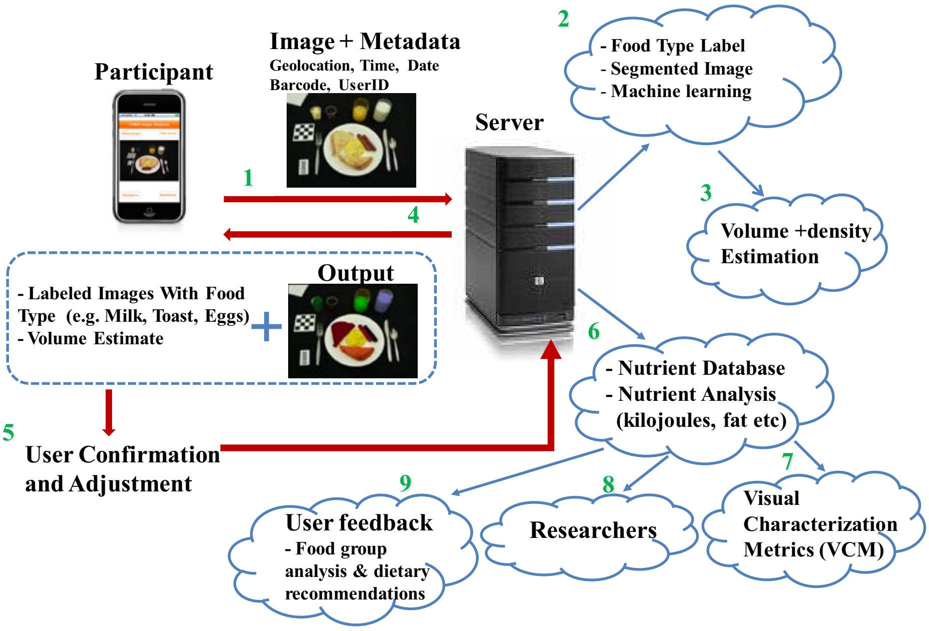


Figure 1. System architecture for the TADA dietary assessment system.

and the volume inside this surface is taken as the volume of the food item.

Martin et al. [12] proposed a system to estimate volume based on a single image of the food. They assumed that the foods captured in the image are approximately bowl-shaped and inferred the volume of the food item from the surface area of the food. They essentially use a predefined linear relationship between surface area and volume. This technique relies on the validity of the bowl shape assumption and is relatively inaccurate for foods that do not satisfy this assumption (spherical fruit for example).

Weiss et al. [13] used a system that required three images of the food item and a calibration grid pattern. This allowed the height of the food item above a reference plane to be estimated and from this height the volume of the food between this 3D surface and the plate was estimated. The requirement to capture three images was a disadvantage of this system as this is considered to be overly burdensome to the participants.

Chen et al. [14] proposed a system that requires only a single image of the food item. The user then manually selects one of a number of predefined geometric models such as a cube, sphere etc. A 3D/2D registration step then automatically aligns the 3D model to the 2D projection of the food item in order to estimate the scale and pose o the 3D model which best matches the 2D projection of the food item. After calibration, the volume of the 3D model at the registered pose and scale is taken as the volume of the food item. This reported estimation errors for this technique were approximately 5% for food items which had a shape that matched one of the pre-defined models and 10% for irregularly shaped objects.

Woo et al. [15] proposed a similar approach which also included irregularly shaped 3D models which consisted of the surface shape of the food item extended towards the surface of the plate to create an irregular prismatic model of the food item. They also allowed manual refinement of the 3D model where the user defined the distance to extend the prismatic model. Chae et al. [16] used 3D model templates for specific common food items including liquids in clear near-cylindrical containers and bread slices. The proposed techniques provided a method for automatically determining the 3D shape of these specific food items (by automatically determining the thickness of the slice of bread on a plate for example).

In this paper we propose a food volume estimation approach similar to that proposed by Weiss et al. However our proposed approach requires only a stereo pair of images to be captured. The proposed approach also includes a novel slice based estimation approach to estimate the volume of a food item from a partial point cloud of the surface of the food item. The proposed approach can be used to estimate the volume of any irregularly shaped food item.

II. The Mobile Phone Based Dietary Assessment System

The food volume estimation procedure proposed in this paper will be utilized in the mobile phone based dietary assessment system that has been developed as part of the Technology Assisted Dietary Assessment (TADA) project. The system architecture for this system is shown in Figure 1. The users of the system are given a mobile phone with a built-in camera, network connectivity, and integrated image analysis and visualization tools to allow their food and beverage intake to be recorded. The process starts with the user sending the food image and contextual metadata (date, time, geo-location, etc.) to the server via the data network (step 1). The user places a fiducial marker in each image that allows the images to be spatially and color calibrated. Food identification (and in the future volume and density estimation) is performed on the server (steps 2 and 3) for finding the nutrient information (step 6). The database contains information on the most commonly consumed foods, their nutrient values, and weights and densities for typical food portions. The food database uses image based features, referred to as Visual Characterization Metrics (VCMs) (step 7), to index the nutrient information. Finally, these results are sent to the researcher for further analysis (step 8) and future developments will incorporate user feedback including food group analysis and dietary recommendations (step 9).

III. 3D Reconstruction of Food Items

In order to estimate the volume of any food object in the image based dietary assessment system, firstly we need to reconstruct the 3D shape of the food. The three dimensional (3D) reconstruction of objects from single/multiple view/s is an on-going research area in the field of computer vision. Humans interpret depth using various visual cues to understand the third dimension of an object, in the 3D world. However, for a given two dimensional (2D) image, we have the ability to visualize the third dimension through information on perspective projection from the images. The interest lies in the process of gathering this 2D image data and processing it to create the 3D structure. Hence, 3D reconstruction involves the use of techniques in computer vision to add the missing dimension to create the 3D space from 2D images.

To describe the process of image acquisition, we will use the pinhole camera model which projects 3D objects onto a 2D image plane. We will then introduce the basic geometry used for the reconstruction of points in 3D space using two different (calibrated) camera viewpoints. Finally, a simple algorithm which can be used to recover the 3D position of such points from their 2D views will be explained.

A. *Pinhole Camera Model*

A point in 3D space $w=[X, Y, Z]^T$ and its projection in a 2D image plane $m=[x, y]^T$ can be represented using the pinhole camera model as shown in Figure 2(a). The relationship of the projection between the two planes can be expressed using the homogeneous coordinate:

$$\tilde{m} = M\tilde{w} \tag{1}$$

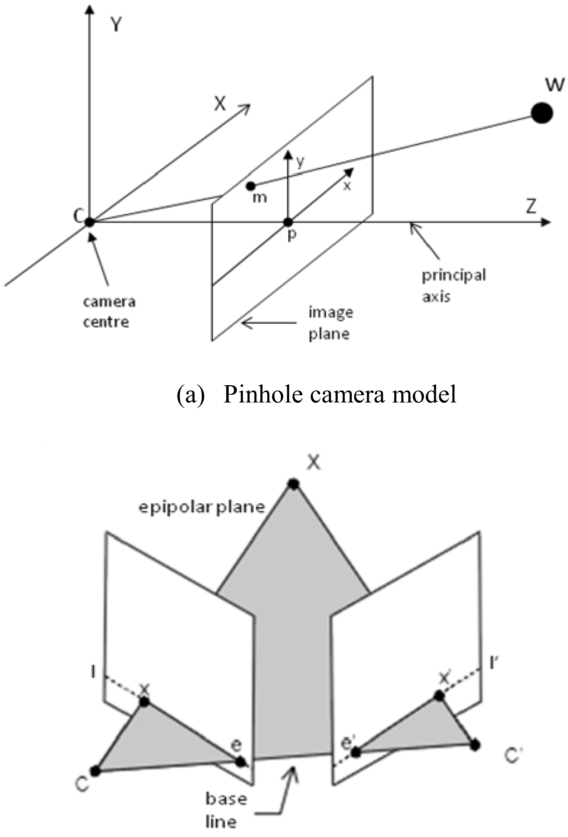
Here M is the perspective projection matrix which can be represented as

$$M = K[R \mid t] \tag{2}$$

K is the camera matrix containing internal parameters, given by

$$K = \begin{bmatrix} \alpha & s & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where $\alpha = fk_x$, $\beta = fk_y$, f is the focal length, k_x and k_v are the distortions along the image plane, s is the skew factor, and (x_0, y_0) is the principal point, i.e. the point where the optical axis intersects with the image plane.



(a) Pinhole camera model

(b) Epipolar geometry

Figure 2. Camera and projection

B. *Epipolar Geometry*

Epipolar geometry refers to the intrinsic projective geometry between two views. It is independent of the structure in the scene and only depends on the cameras' intrinsic parameters and relative positions. Suppose a point X in 3D space is captured by two views, at x in the left view and x' in the right view as shown in Figure 2(b). Then image points x and x', space point X, and camera centers C and C' are coplanar. This plane is called the epipolar plane and can be determined by the ray back-projecting from x to X and the base line joining the camera centers. The points e and e' where the base line of the cameras intersects with the image planes are called epipoles. Epipoles are actually the images of the camera centers and all epipolar lines must pass through them. In the context of stereo matching algorithms for depth estimation, the search for x' that corresponds to x can be constrained to the epipolar line instead of the whole image. Further details of epipolar geometry and its terminology can be found in [17].

C. *Fundamental Matrix*

The fundamental matrix F is a 3x3 matrix with rank 2 and satisfies the condition that, for any pair of corresponding points x and x' in the two images, the following equality holds:

$$x'TFx = 0 \tag{4}$$

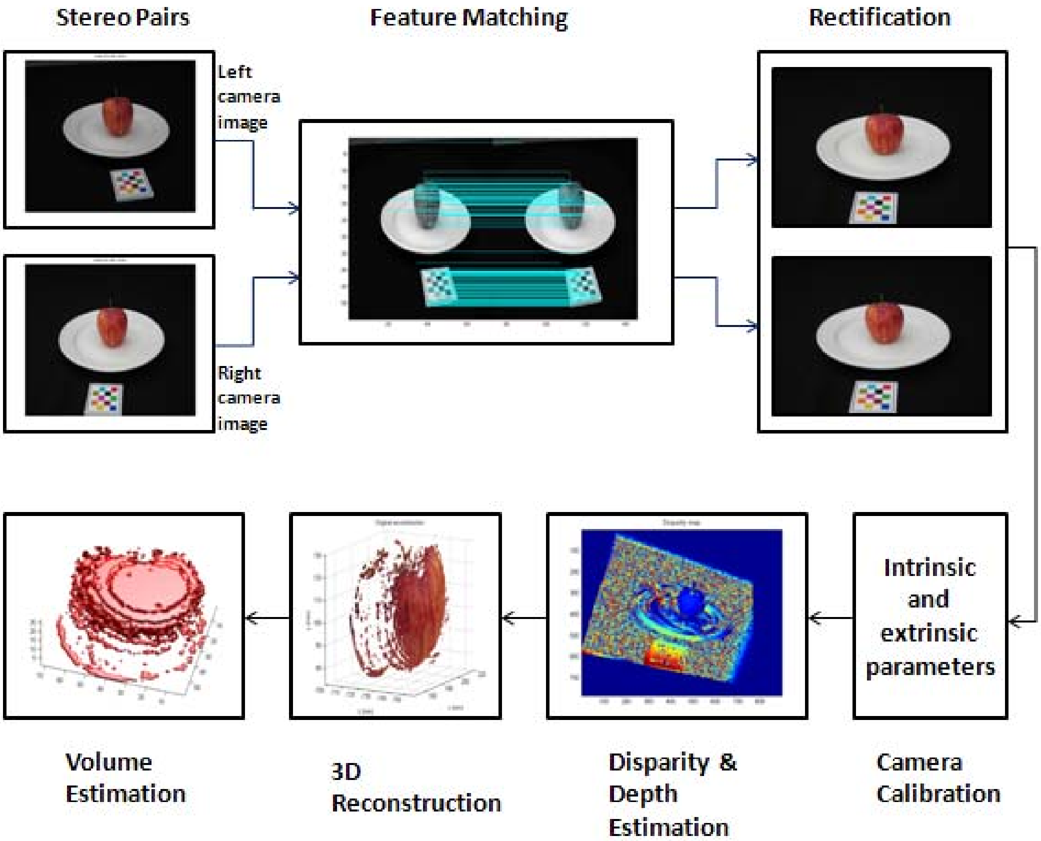


Figure 3. Flow chart of the proposed volume estimation approach

According to epipolar geometry, x' lies on the epipolar line $l' = Fx$. Hence $x'^T l' = 0$ and then $x'^T Fx = 0$. We use the strong camera calibration method where the fundamental matrix can be computed using the calibrated camera projection matrices P and P'. Given P and P', the fundamental matrix F can be represented by

$$F = [e]_{\times} P' P^{+} \qquad (5)$$

where P^{+} is the pseudo-inverse of P, i.e. $PP^{+} = I$, and $[a]_{\times}$ is defined as in (6) when $a = (a_1, a_2, a_3)^T$.

$$[a]_{\times} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \qquad (6)$$

The fundamental matrix, F can also be computed using corresponding points from two views. Given $x = (x, y, 1)^T$ and $x' = (x', y', 1)^T$, the equation (4) can be rewritten as

$$(x'x, x'y, x', y'x, y'y, y', x, y, 1) f = 0 \qquad (7)$$

where $f = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^T$ and f_{ij} are the corresponding elements of F. With a set of n point matches, a set of linear equations are obtained as follows

$$Af = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ . & & & & & & & & . \\ . & & & & & & & & . \\ . & & & & & & & & . \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} f = 0 \qquad (8)$$

F can be computed by solving this set of equations and will be used in the rectification process, which is essential for depth map calculation.

IV. Proposed Food Volume Estimation Approach

Accurate estimation of food portion size from the captured views is one of the challenging problems for a dietary assessment system._We have developed a novel slice based volume estimation method, which requires a set of two 2D images to be taken from the left side and the right side of the food item located on the user's plate, to automatically estimate the portion size of a variety of foods. The food volume estimation procedure consists of 6 steps: feature matching, rectification, camera calibration, disparity & depth map generation, 3D structure reconstruction and 3D volume estimation. Figure 3 shows a flowchart of the proposed volume estimation approach.

A. *Feature Matching*

Feature matching identifies the corresponding feature points between the stereo pairs and is used for image rectification. In this paper, we apply the SIFT algorithm [19] on the stereo pair in order to determine the corresponding feature points. The SIFT algorithm extracts the local features of the objects at particular interest points. The extracted features are not only invariant to image scale and rotation but also robust to changes in illumination, noise, and minor changes in viewpoint. These properties make the extracted features highly distinctive allowing correct object identification with low probability of mismatch. As the rectification process [17] requires at least 8 correct matches, we randomly select 16 matched feature points (from the checker board pattern and segmented food region) and the rest are removed. An example of the matched points between the stereo images of an apple is shown in Section V.

B. Stereo Rectification

The fundamental matrix can reduce the search of corresponding points from the whole image to the epipolar line. In practice, a pre-processing procedure known as rectification is usually required in most stereo-matching based disparity and depth estimation algorithms. In our rectification process, a pair of images taken from different viewpoints are transformed and re-sampled to produce a pair of rectified images in which the epipolar lines are all parallel with the *x*-axis and are the same in both views. Consequently, the disparity only exists in the *x*-direction. In fact, if the epipolar line is parallel with the *x*-axis after transformation, the epipole will be transferred to the infinite point with the 2D homogeneous image coordinate equal to $(1, 0, 0)^T$ [17]. Suppose, in the partially transformed left view, a point of interest u_0, i.e. the center of the image, is the origin and the epipole e with a 2D homogenous image coordinate $(f, 0, 1)^T$ lies on the *x*-axis, then e can be transferred to the infinite point with the following transformation

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{bmatrix} \qquad (9)$$

Finally, the required transformation for rectification of the original left view is given by the product $H = GRT$ where T is the translation taking u_0 to the origin, R is a rotation around the origin taking e to the *x*-axis and G is the transformation which maps e to the infinite point. As for the corresponding rectification transformation H' for the original right view, it can first be computed using a similar method and then optimized by minimizing the sum of squared distance $\sum d(Hx_i, H'x_i')^2$, where x_i and x_i' are corresponding points and $d(x_i, x_j)$ is the distance between the points x_i and x_j. A complete description of image rectification with detailed analysis can be found in [17] and [18]. The rectification results for a pair of apple images are shown in the experimental results (Section V).

C. Camera Calibration

Camera calibration is a necessary step in 3D reconstruction to extract metric information from 2D images [19]. In our system, we consider a very simple protocol, which involves the use of a calibrated fiducial marker. It consists of a checker board (color) with known dimensions placed in the field of view of the camera as shown in Figure 3. This allows us to identify the scale and pose of the food item to be estimated and also allows color correction of the images.

The camera parameters consist of intrinsic parameters (focal length, principal point, distortion and skew) and extrinsic parameters (camera orientation and translation). In our food volume estimation process, we use the fiducial marker as a reference to measure the amount of the food present in the plate.

In our experiments, we compute the camera intrinsic parameters using Camera Calibration Toolbox [22] from the California Institute of Technology. Such a tool will produce an intrinsic matrix, KK, of the form:

$$KK = \begin{bmatrix} fc(1) & alpha\_c * kc(1) & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 0 \end{bmatrix} \qquad (10)$$

where fc is a 2×1 vector which contains the focal length in pixels, cc is a 2×1 vector which stores the principal point coordinate, $alpha\_c$ is the skew coefficient (the angle between the x and y axis) and kc is a 5×1 vector, represents the image distortion coefficients.

D. Disparity and Depth Map Generation

A disparity map is a depth map where the depth information is derived from offset images of the same scene. Depth maps can be generated using various other methods, such as time-of-flight (sonic, infrared, laser). Although these active methods can often produce far more accurate maps at short distances, the passive method has its benefits, including applicability at long distances. The depth information from the active approaches [21] are either not accurate enough or based on strong assumptions i.e. shape priors or static objects which make them not suitable for multi-view reconstruction applications. In contrast, passive techniques rely solely on images captured by cameras and depth from these approaches (e.g., stereo matching) is straightforward and reliable.

The simple principle of depth estimation from stereo matching is illustrated in Figure 4, where *B* is the length of the baseline joining the camera centers, *f* is the camera focal length and *disparity* = |*x* – *x*'|. Then

$$depth = \frac{Bf}{disparity} \qquad (11)$$

which means depth is inversely proportional to disparity and disparity is often treated as synonymous with inverse depth. The disparity map generated using normalized cross correlation for the apple image is shown in the experimental results (Section V).

E. 3D Reconstruction

Once the disparity map is estimated, we can convert these disparities into depths using equation (11). With the depth map and knowledge of the intrinsic parameters of

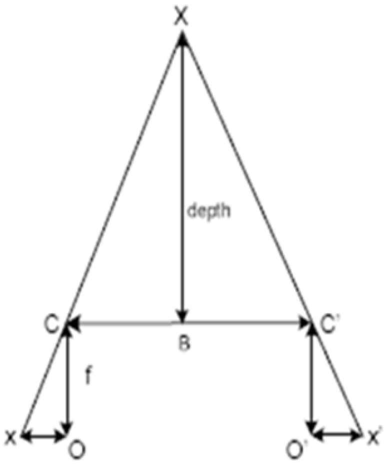


Figure 4. Depth vs. disparity

the camera, we are able to recover the 3D location of all points in the image plane[17, 19].

This camera matrix relates 3D world coordinates to homogenized camera coordinates via

$$\left[X_{camera} \quad Y_{camera} \quad 1\right]^T = KK.\left[X_{world} \quad Y_{world} \quad Z_{world}\right]^T \quad (12)$$

If we know the intrinsic matrix, we can back project each image pixel into a 3D ray that describes all the world points that could have been projected onto that pixel on the image plane. However, the distance of that point to the camera is unknown. This can be recovered by the disparity measurements of the stereo depth map as

$$Z_{world} = focal\_length * \frac{stereo\_baseline}{disparity} \quad (13)$$

Here the pixel disparities are unitless; hence they cannot be used directly in this equation. Also, if the stereo baseline (the distance between the two cameras) is not well-known, then it introduces more unknowns. Thus we transform this equation into the general form

$$Z_{world} = p + \frac{q}{disparity} \quad (14)$$

There are two unknowns *p* and *q* in this equation, thus we solve this using least square method by collecting a few corresponding depth and disparity values from the fiducial marker and we use them as tie points.

In our reconstructed 3D scene, we see that only the front view is reconstructed leaving the other part unconstructed. The reason for this is the back view was occluded and we cannot recover the occluded scene thus we use a symmetric reconstruction approach to recover the occluded regions. Symmetry is a universal concept in nature, science, and art. In the physical world, geometric symmetries and structural regularity occur at all scales, from crystal lattices and carbon nano-structures to the human body, architectural artifacts, and the formation of galaxies [23]. Naturally many food objects are symmetric in the real world e.g. apple, orange, pineapple etc. If we can reconstruct half of the food object, we can often easily reconstruct the other half using symmetry.

F. Volume Estimation

In the previous step, the 3D structure of the food item is recovered as a 3D point cloud (a set of vertices in a three-dimensional coordinate system). Although the point cloud can be viewed and inspected directly, they are not generally usable in most 3D applications, and therefore are usually converted to polygons or triangular mesh models, NURBS surface models, or CAD models through a process commonly referred to as surface reconstruction [23].

In our system, we convert the reconstructed 3D point cloud into a series of slices. We accomplish this by dividing the 3D point cloud into several slices. These slices contain exactly the same information as the initial point cloud - the (*x, y, z*) coordinates of the points. The points of each slice are co-planar, so we can process each slice as a 2D set of points, instead of a 3D object. The new slices in the 3D body help to access the local information of a particular slice and the information between adjacent slices efficiently allowing the reconstruction of global structure and the shape of the food object.

We divide the point cloud into slices along the *z*-axis, which represents the depth information. We convert each slice into binary data to perform some morphological operations to enhance the object shape. The holes inside the slices are filled and the slices are segmented to extract the locations of boundary points. On each slice, we fit a circle/ellipse (based on food shape) and we find the radius (for circle), length of major and minor axis (for ellipse) to estimate the area under the food region. Simply stated, the volume estimation process can be performed in two steps: First, form 2D slices from the 3D point cloud and second, find the total volume (V_T) of the food object by summing the individual volumes (V_k) for each slice. If *S* is the total number of slices, then we can compute the total volume as follows

$$V_T = \sum_{k=1}^{S} V_k \quad (15)$$

V. Experimental Results

To demonstrate accuracy of the proposed food volume estimation approach, we performed validation experiments on 6 fruit items namely: apple, orange, pear, banana, pineapple and kiwi-fruit as shown in Figure 5. Each fruit item was placed on a plate on a table and a pair (left and right) of images for each category of fruit were captured using an iPhone 4S. In our settings, each image contains a checker board pattern to estimate the scale and pose of that image and the length of each side of the square is 11 mm. The image pair is first fed into the SIFT algorithm and then we follow the steps described in the previous section (Section IV) to recover the 3D structure of the food as a series of 2D slices. The step by step result of the proposed approach for an apple is shown in Figure 6. In our volume estimation process, we fit a circle on the reconstructed slice data of the apple, orange and pear, and we use an ellipse for the pineapple, banana and kiwi–fruit. The volume of each slice is calculated first and then the total

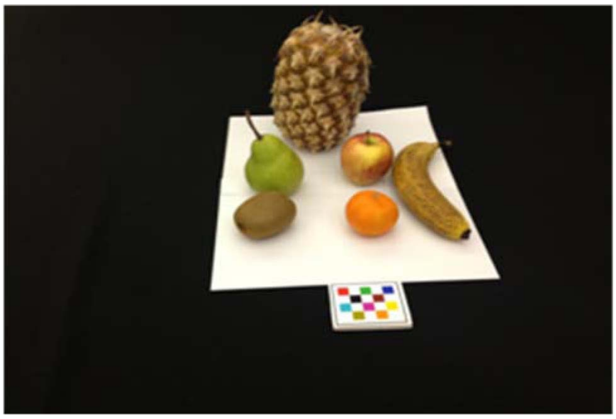


Figure 5. Food items used in our experiments

volume of the fruit can be calculated using equation (15). The ground truth volume of each fruit was measured by the_water displacement method. The estimated volumes of the fruit items using the proposed approach are shown in Table I and they are compared with their respective ground truths. The average error of our proposed approach is 7.7 %, which is comparable to the accuracies of other more complex approaches.

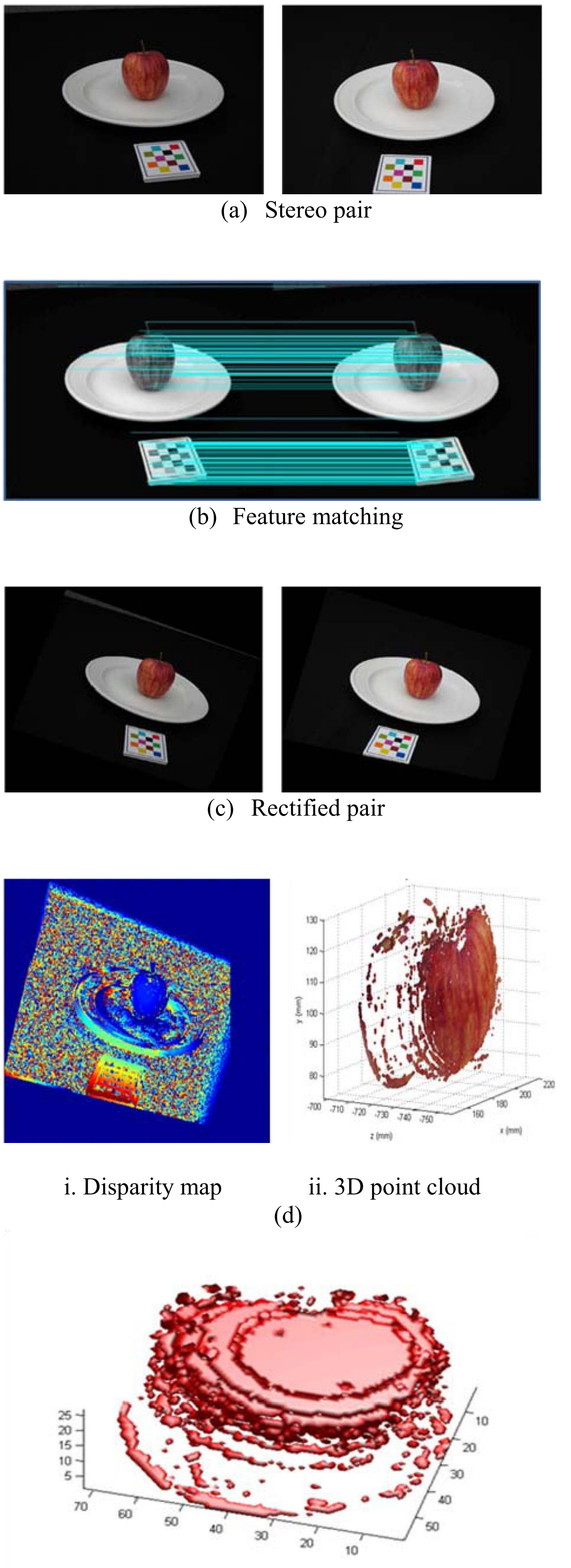
Our experiments show that the less textured food items may lead to erroneous 3D volume estimation, which can be improved by using an algorithm having better dense matching. Non-uniform lighting conditions also cause inaccuracy in the volume estimates.

TABLE I. VOLUME ESTIMATION ERROR (%)

Food item	Ground truth (ml)	Estimated (ml)	Error (%)
Apple	185	181.5	2.7
Orange	118	112.6	4.6
Pear	230	195.1	15.1
Banana	155	146.5	5.4
Pineapple	1547	1404.2	9.4
Kiwifruit	120	108.8	9.3
Average error			7.7

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a food volume estimation approach by geometrically reconstructing the objects from a pair of stereo images. We consider a free hand approach where the user can capture pictures with minimal restrictions on the position and distance between the cameras. Our image acquisition step requires the inclusion of a fiducial marker in each food image. We obtain the camera parameters and then we back project the image plane into 3D world coordinates. We tested the accuracy of the proposed approach by performing experiments on 6 types of popular fruits. The experimental results showed that our approach has an average error of less than 10%. This indicates that the proposed approach can provide an accurate estimate of the volume of typical food items in a passive manner without the need for manual fitting of 3D models to the food items.



(a) Stereo pair

(b) Feature matching

(c) Rectified pair

i. Disparity map ii. 3D point cloud

(d)

(e) 3D slices (A series of 2D slices)

Figure 6. The step-by-step results of the proposed approach for the apple images.

In future work, we intend to use a 3D stereoscopic camera in a smart-phone to obtain the 3D structure of the food and then estimate the volume of that reconstructed 3D shape. We will then compare the accuracy of the freehand food volume estimation approach with the approach using the 3D stereoscopic camera.

REFERENCES

[1] World Health Organization, World Health Assembly resolution WHA57.17 – "Global strategy on diet, physical activity and health." 2003, World Health Organization: Geneva.

[2] World Health Organization, WHO Technical Report Series 916 – "Diet, nutrition and the prevention of chronic diseases," R.o.a.J.W.F.E. Consultation, Editor. 2003: Geneva.

[3] B. M. Margetts, and M. Nelson, "Overview of the principles of nutritional epidemiology," in Design concepts in nutritional epidemiology, B.M. Margetts and M. Nelson, Editors. 1997, Oxford University Press: Oxford ; New York. p. 3-38.

[4] Australian Bureau of Statistics, "Cafes, Restaurants and Catering Services, Australia, 2006-07." 2008, Australian Bureau of Statistics, Canberra.

[5] Australian Government Department of Agriculture Fisheries and Forestry, "FOODmap. A comparative analysis of Australian food distribution channels," Commonwealth of Australia, Editor. 2006.

[6] J. Dixon, et al., "The health equity dimensions of urban food systems," J Urban Health, 2007. 84(3 Suppl): p. i118-29.

[7] C. Burns, et al., "Foods prepared outside the home: association with selected nutrients and body mass index in adult Australians," Public Health Nutr, 2002. 5(3): p. 441-8.

[8] C. J. Boushey, et al., "Use of technology in children's dietary assessment," European Journal of Clinical Nutrition, 2009. 63: p. S50-S57.

[9] J. T. Tufano and B.T. Karras, "Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends," J Med Internet Res, 2005. 7(5): p. e58.

[10] J. Shang, et al., "A pervasive Dietary Data Recording System," Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on , pp. 307-309, 21-25 March 2011.

[11] J. Shang, et al., "Dietary intake assessment using integrated sensors and software." Proc. SPIE 8304, Multimedia on Mobile Devices and Multimedia Content Access: Algorithms and Systems VI, 830403, 9 February 2012.

[12] Martin, C.K.; Kaya, S.; Gunturk, B.K.; , "Quantification of food intake using food image analysis," Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE , pp.6869-6872, 3-6 Sept. 2009.

[13] Weiss, R., Stumbo, P.J., Divakaran, A., "Automatic Food Documentation and Volume Computation Using Digital Imaging and Electronic Transmission," Journal of the American Dietetic Association, Volume 110, Issue 1, January 2010, Pages 42-44.

[14] Chen, H.-C., et al., "3D/2D model-to-image registration for quantitative dietary assessment," Bioengineering Conference (NEBEC), 2012 38th Annual Northeast , pp.95-96, 16-18 March 2012.

[15] Woo I., et al., "Automatic portion estimation and visual refinement in mobile dietary assessment," Proc. SPIE 7533, Computational Imaging VIII, 75330O, January 27, 2010.

[16] Chae, J., et al., "Volume estimation using food specific shape templates in mobile image-based dietary assessment." Proc. SPIE 7873, Computational Imaging IX, 78730K, February 07, 2011.

[17] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press, UK, 2003

[18] A. Fusiello, E. Trucco and A. Verri, A Compact Algorithm for Rectification of Stereo Pairs. In Machine Vision and Applications, 12 (1): 16-22, 2000.

[19] Zhang, Z., "A flexible new technique for camera calibration," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.22, no.11, pp. 1330- 1334, Nov 2000

[20] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Int. J. Comput. Vision,60(2), 91-110, Int. J. Comput. Vision, 2004

[21] S.B. Gokturk, H. Yalcin, and C. Bamji. "A time-of-flight depth sensor- system description, issues and solutions." In CVPRW, 8, 24, page 35, june 2004.

[22] http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html

[23] N. J. Mitra and M. Pauly and M. Wand and D. Ceylan, "Symmetry in 3D Geometry: Extraction and Applications", 33rd Annual Conference of the European Union for Computer Graphics (Eurographics 2012), Cagliari, Italy, May 13-18.