DIMENSIONAL MEASUREMENT OF OBJECTS IN SINGLE IMAGES INDEPENDENT FROM RESTRICTIVE CAMERA PARAMETERS

by

Yaofeng Yue

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science in Electrical Engineering

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Yaofeng Yue

It was defended on

March 24th, 2010

and approved by

Ching-Chung Li, Ph.D., Professor, Electrical and Computer Engineering Department

Robert J. Sclabassi, Ph. D., M.D., Professor Emeritus, Neurological Surgery, Electrical and Computer Engineering Department

John D. Fernstrom, Ph. D., Professor, Psychiatry and Pharmacology

Zhi-Hong Mao, PhD, Assistant Professor, Electrical and Computer Engineering Department

Thesis Advisor: Mingui Sun, Ph. D., Professor, Neurological Surgery, Electrical and Computer Engineering Department

DIMENSIONAL MEASUREMENT OF OBJECTS IN SINGLE IMAGES INDEPENDENT FROM RESTRICTIVE CAMERA PARAMETERS

Yaofeng Yue, M.S

University of Pittsburgh, 2010

Recent advances in microelectronics have produced new generations of digital cameras with variable focal lengths and pixel sizes which facilitate automatic and high-quality imaging. However, without knowing the values of these critical camera parameters, it is difficult to measure objects in images using existing algorithms. This work investigates this important problem aiming at dimensional measurements (e.g., diameter, length, width and height) of regularly shaped physical objects in a single 2-D image free from restrictive camera parameters. Traditionally, such measurements usually require determinations of the poses of a certain reference feature, i.e., the location and orientation of the feature relative to the camera, in order to establish a geometric model for the dimensional calculation. Points or lines associated with certain shapes (including triangles and rectangles) are often used as reference features for the pose estimation. However, with only a single image as the input, these methods assume the availability of 3-D spatial relationships of the points or lines, which limits the applications of these methods to practical problems where this knowledge is unavailable or difficult to estimate, such as in the problem of image-based food portion size estimation in dietary assessment. In addition to points and lines, the circle has also been used as a reference feature because it has a single elliptic perspective projection in images. However, almost all the existing approaches treat the parameters of focal length and pixel size as the necessary prior information. Here, we propose a new approach to dimensional estimation based on single image input using the circular reference feature and a pin-hole model without considering camera distortion. Without knowing

the focal length and pixel size, our approach provides a closed-form solution for the orientation estimation of the circular feature. With additional information provided, such as the size of the circular reference feature, analytical solutions are provided for physical length estimation between an arbitrary pair of points on the reference plane. Studies using both synthetic and actual objects have been conducted to evaluate this new method, which exhibited satisfactory results. This method has also been applied to the measurement of food dimensions based on digital pictures of foods in circular dining plates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis was motivated by the work of food volume estimation in a dietary assessment study supported by the National Institutes of Health grant U01 HL91736. The preliminary work using a circular feature (e.g a spotlight pattern shining on a food or a flat surface) as a physical reference to calculate food portion size using one single image was conducted by Prof. Wenyan Jia and her colleagues. The implementation of this method was reported in the literature. In this method, one needs to know the focal length, pixel size and other camera parameters in order to take food dimension measurements. A special advantage of this method over other methods lies in that the circular feature (e.g. a spotlight, a plate or a bowl) serves an excellent reference to be used in dietary study. However, in many cases, the food images acquired in the real world do not have the focal length and pixel size information. It was my advisor, Prof. Mingui Sun who hypothesized that certain 3-D measurements could be conducted using a circular reference feature without knowing restrictive parameters of the camera. This hypothesis stimulated my further effort to bring my own contribution to the emerging research on obesity using the computer vision technology. Specifically, this thesis will prove the validity of this hypothesis.

More importantly, I would like to take this opportunity to direct my warmest thanks to Prof. Mingui Sun and Prof. Zhi-Hong Mao for providing me the precious opportunity in continuing

study at the University of Pittsburgh. I would also like to thank Prof. Sclabassi, Prof. Fernstrom, and Prof. Jia and all my labmates for their valuable discussions. My thanks also go to my wife, Qilu Zeng, for her understanding and constant support during the course of this research.

1.0 INTRODUCTION

Dimensional measurement of physical objects in a single 2-D image is a fundamental problem in machine vision. It has wide applications in many fields, such as robotics, industrial automation, security and surveillance. Nowadays, the digital camera as a central component of the computer vision system has advanced to a highly intelligent level. For example, almost all new-generation digital cameras have the feature of auto-focusing. Among them, many are featured with automatic change in image resolution controlled by certain variables. It is expected this resolution change capability to become increasingly popular among new commercial products in the near future. With these technological advances in place, it is necessary for the engineers in the computer vision field to revisit the fundamental problem of modeling the process of imaging involving a camera and a physical object. Undoubtedly, the unknown focal length and pixel size impose new problems in applying the traditional approaches to dimensional measurement of regularly shaped objects using one single image. How can we obtain more information from the image without knowing focal length and pixel size? This work attempts to answer this important question.

To conduct dimensional measurement using one single image, we normally need one or several features as the reference and require that the object under measurement be placed on the same plane of the reference. This problem is essentially equivalent to estimating the relative location and orientation of the reference to the camera (i.e., pose estimation), which is necessary

to build geometric models to estimate dimensions on the reference plane (details will be provided in section 2.3). The location and orientation estimation is usually modeled as the calibration of camera parameters, which establishes the perspective transformation between the 3-D object and its 2-D image [Sonka et al. 2007; Haralick 1980]. Within this process, the selected reference feature plays the key role in establishing the mathematic model and finding the solution. In current literature, features used as references to estimate camera parameters can be generally classified into two groups: points (including lines) and parameterized curves. The use of points is known to be the "perspective-n-point" problem (PnP) [Fischler and Bolles 1981], Given the relative spatial locations of *n* control points, and given the angle to every pair of control points from an additional point called the Center of Perspective (CP), the solution to the "perspective-n-point" problem provides the lengths of the line segments ("legs") joining the CP to each of the control points. Within the PnP problem, the studies using three-points (or a triangle) or four-points (or a rectangle) as reference features were most popular. Fischler and Bolles [1981] pointed out that there are four possible solutions to the P3P problem but only one unique solution to the P4P problem when the four points are on the same plane. The P3P problem was studied further from different perspectives [Wolfe and Jones, 1986; Linnainmaa et al. 1988; Wolfe et al. 1991; Menthon and Davis 1992; Gao et al. 2003]. In addition to the P3P problem, it has been shown that the camera parameters can be estimated using the perspective projection of a rectangle of know size and position, which is equivalent to the P4P problem [Haralick 1980; Haralick 1989; Horaud et al. 1989; Abidi and Chandra, 1995]. The quadrangular feature, a more generalized P4P problem, has also been studied [Abidi and Chandra, 1995]. As noticed, the successful solutions to the P3P and P4P problems require the 3-D locations of each pair of the points to be known, which may impose too many restrictions in certain practical applications. In

contrast to the usage of points, choosing curves as references in the estimation of camera parameters may help us overcome this limitation. Haralick [1984] studied the estimation of camera parameters for a general parameterized quadratic curve. However, as a particular instance of the quadratic curve, the circular feature used as a reference gained more popularity later and was studied in robotics by Magee and Agarwal [1984] and Kabuka et al. [1987], where only the focal length and size of the circular pattern need to be known. Because circular objects and patterns are very common, and their perspective projections are always ellipses [Narayan 1961], the circular feature as a potential reference in the estimation of camera parameters serves a good choice in certain applications. Furthermore, Mulgaonkar [1984] addressed the viewpoint of a circular plane using an iterative method. Marimont [1986] and Sawhney et al. [1990] presented a closed-form solution for the circular feature based on linear algebra. Shin and Ahmad [1989] and Safaee-Rad et al. [1992] provided closed-form solutions derived from geometric models.

In this work, we study the dimensional measurement problem based on the single image input using the circular feature as a reference, and apply the result of this study to dietary assessment. We aim to eliminate the requirements of the focal length and pixel size, because currently most images are captured by modern digital cameras with variable focal lengths and/or pixel sizes. Almost all above mentioned approaches, on the other hand, treated the focal length and pixel size as necessary information and hence cannot be applied directly to this new type of digital images.

In order to cope with the problem of unknown focal length and pixel size, we propose a new approach which provides closed-form solutions to the orientation of the circular feature. With additional information given, such as the diameter of the reference feature, analytical solutions for the dimensional measurements of an object on the reference plane are also provided.

This approach is based on the simplest pin-hole model, and currently does not consider the effect of camera distortion. Our derivation process of this approach provides us with the closed-form expressions of a number of useful intermediate parameters, including the ratio of the focal length respected to the pixel size, the orientation of the reference plane, and the distance from the optical center to the reference plane (details described in section 2.1) Experiments using both synthetic and real objects were conducted to test and analyze this approach, which exhibited satisfactory results in dimensional measurements of regularly shaped objects. However, for the three intermediate parameters, the results were not as satisfactory as those in the dimensional measurements. For example, the measured orientation had an error ranging up to between ten and twenty percent. Although our simulation showed that the errors for the intermediate parameters resulted from the ideal pin-hole model, the correction of distortion due to the model to improve the estimation of these intermediate parameters requires further research. This research may also provide an answer to the question why relatively large errors in the intermediate variables do not carry to the final result of dimensional measurements. Therefore, more extensive exploration is required for a thorough understanding of the mechanisms in the estimation process.

As a specific application, we utilized our method to measure food dimensions which is important in the study of the rising health problem of obesity. Traditionally, dietary assessment is conducted based on self-reporting of food intake. A food diary, which lists food names and portion sizes, or a dietary recall, which is conducted for a period of time (e.g., 24 hours), is often used in clinical practice. However, this method is known to be inaccurate [Trabulsi and Schoeller 2001]. In recent years, digital cameras and cell phones have advanced rapidly. It has been reported that these personal imaging and communication devices can be used for dietary

assessment with much improved accuracy [Engelen et al. 2009 and Sun et al. 2010]. In order to estimate food intake from digital pictures, the volume (or the portion size) of each food photographed must be determined. However, given a single image, the volumetric information cannot be obtained without a dimensional reference and an assumption of food shape. As discussed above, the approaches of using N points as references are limited in that the spatial relationships of each pair of the points need to be known. Therefore, it is impractical and inconvenient for a human subject to bring a reference object (e.g. a monochrome or color checkerboard) to be placed next to the food before taking a picture. On the other hand, since most foods are served using circular plate(s) or bowl(s), these objects can be used as references. Therefore, the approach of using circular reference feature becomes especially appropriate for this specific application.

2.0 METHOD

2.1 THE MODEL

The well known pin-hole camera model provides a perspective projection from the object plane to the image plane. Figure 1 illustrates the geometric projective relationship between a circular reference object on the object plane and its projected feature, an ellipse, in the image. The vertex point O, the circular reference object, and the radial lines from point O passing through the boundary of the circular object form a cone. In this section, we will establish a mathematical model for this cone. Then, we will transform the conic region to a new coordinate system under which the original base of the cone will be perpendicular to the new z axis so that this base can be expressed by the standard circular equation. Since several coefficients of the new expression are known to be zero, a system of equations can be obtained which will be solved for the desired parameters.

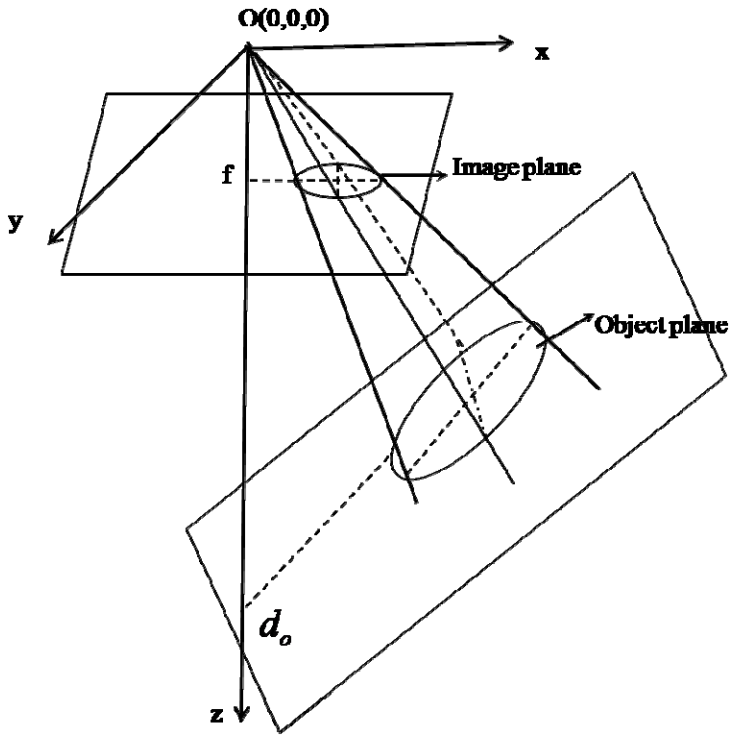

Figure 1.Schematic representation of the optical center, image and object.

Since the perspective projection of a circular reference feature is always an ellipse [Narayan 1961], we start with the well-know mathematical description of the ellipse. The general form of the ellipse is given by

$$a'x^2 + 2h'xy + b'y^2 + 2g'x + 2f'y + d' = 0 \qquad (1)$$

Where, a', h', b', g', f' and d' are coefficients, $d' \neq 0$, and $d'\left(h'^2 - ab\right) > 0$

The six unknown coefficients need to be determined in order to specify an ellipse. Many methods are available to fit an ellipse using at least six points [Fitzgibbon et al. 1999]. However, since we do not know the focal length and pixel size, it is impossible to explicitly describe the ellipse in the input image with respect to its physical parameters. The idea of starting to work with the ellipse equation is to use the linearity of the first five unknown coefficients to obtain new representations. To do this, the ellipse information which is available in the pixel coordinate

in the image needs to be converted to the 2D real-world coordinates as a function of unknown focal length and pixel size. Here, we first select five points on the boundary of the ellipse on the image. Let (x_i, y_i), $i = 1,\cdots,5$, be the real coordinates of these five points; and let (x_{pi}, y_{pi}), $i = 1,\cdots,5$, be their pixel coordinates in the image. If we define (c_1, c_2) as the pixel coordinate of the center of the image and p_x as the pixel size, then the two difference coordinates are related by:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} x_{pi} - c_1 \\ y_{pi} - c_2 \end{bmatrix} \cdot p_x \qquad (2)$$

In addition, if we define

$$A = \begin{bmatrix} (x_{p1}-c_1)^2 & 2(x_{p1}-c_1)(y_{p1}-c_2) & (y_{p1}-c_2)^2 & 2(x_{p1}-c_1) & 2(y_{p1}-c_2) \\ (x_{p2}-c_1)^2 & 2(x_{p2}-c_1)(y_{p2}-c_2) & (y_{p2}-c_2)^2 & 2(x_{p2}-c_1) & 2(y_{p2}-c_2) \\ (x_{p3}-c_1)^2 & 2(x_{p3}-c_1)(y_{p3}-c_2) & (y_{p3}-c_2)^2 & 2(x_{p3}-c_1) & 2(y_{p3}-c_2) \\ (x_{p4}-c_1)^2 & 2(x_{p4}-c_1)(y_{p4}-c_2) & (y_{p4}-c_2)^2 & 2(x_{p4}-c_1) & 2(y_{p4}-c_2) \\ (x_{p5}-c_1)^2 & 2(x_{p5}-c_1)(y_{p5}-c_2) & (y_{p5}-c_2)^2 & 2(x_{p5}-c_1) & 2(y_{p5}-c_2) \end{bmatrix} \qquad (3)$$

Then the ellipse equations for these five points can be written in the following matrix form

$$A \cdot \begin{bmatrix} p_x^2 & & & & \\ & p_x^2 & & & \\ & & p_x^2 & & \\ & & & p_x & \\ & & & & p_x \end{bmatrix} \cdot \begin{bmatrix} a' \\ h' \\ b' \\ g' \\ f' \end{bmatrix} = -d' \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad (4)$$

Thus, we can obtain the unknown parameter and coefficients as follows

$$\begin{bmatrix} a' \\ h' \\ b' \\ g' \\ f' \end{bmatrix} = \begin{bmatrix} p_x^2 & & & & \\ & p_x^2 & & & \\ & & p_x^2 & & \\ & & & p_x & \\ & & & & p_x \end{bmatrix}^{-1\cdot} \cdot A^{-1} \cdot -\underline{1} \cdot d' \qquad (5)$$

The right hand side has only one unknown coefficient d'. If we further denote the known vector,

$$A^{-1} \cdot -\underline{1} = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 \end{bmatrix}^T \qquad (6)$$

then the five coefficients can be solved as a function of the unknown pixel size and d',

$$\begin{bmatrix} a' \\ h' \\ b' \\ g' \\ f' \end{bmatrix} = \begin{bmatrix} t_1 p_x^{-2} \\ t_2 p_x^{-2} \\ t_3 p_x^{-2} \\ t_4 p_x^{-1} \\ t_5 p_x^{-1} \end{bmatrix} \cdot d' \qquad (7)$$

Now we consider the pinhole model between the reference circular feature and its perspective projection on the 2-D image. Figure 1 shows the schematic diagram of the camera center, image and the object (i.e circular reference feature). In Figure 1, O is the optical center of the camera, f is the unknown focal length, d_o is the z coordinate of the intersection point of the object plane and z axis. The optical center and the circular feature composes a cone, which interests with the image plane. The intersection area is an ellipse which is exactly the perspective projection of the circular feature [Narayan 1961].

In the coordinate system xyz (Fig. 1), we would like to determine the cone equation which can be defined by a base of the elliptic feature on the image plane and the lines passing through the vertex point O and the boundary of the ellipse. The equation of the lines passing the point O can be described as,

$$\frac{x}{v_x} = \frac{y}{v_y} = \frac{z}{v_z}, \quad (8)$$

where (v_x, v_y, v_z) indicates any direction vector.

The intersection points of above lines with the image plane $z = f$ are $\left(\frac{v_x}{v_y} f, \frac{v_y}{v_z} f, f\right)$. Since these points need to be on the boundary of the ellipse, substituting these points into (1), we obtain

$$a'\frac{v_x}{v_z} f^2 + 2h'\frac{v_x v_y}{v_z^2} f^2 + b'\frac{v_y^2}{v_z^2} f^2 + 2g'\frac{v_x}{v_z} f + 2f'\frac{v_y}{v_z} f + d' = 0 \quad (9)$$

.

Combining (8) and (9), the cone equation can be obtained,

$$a' f^2 x^2 + 2h' f^2 xy + b' f^2 y^2 + 2g' fxz + 2f' fyz + d' z^2 = 0 \quad (10)$$

Or in the matrix form,

$$\begin{bmatrix} x & y & z \end{bmatrix} \cdot Q \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0, \text{ where } Q = \begin{bmatrix} a' f^2 & h' f^2 & g' f \\ h' f^2 & b' f^2 & f' f \\ g' f & f' f & d' \end{bmatrix} \quad (11)$$

With the equation of the cone surface established, we now determine the expression of the intersecting surface of the cone and object plane. We assume a normalized vector (l, m, n) to be the orientation of the object plane. Because it intersects the z axis at point $(0,0,d_o)$, its equation can be established as follows:

$$lx + my + nz = nd_o \qquad (12)$$

where

$$l^2 + m^2 + n^2 = 1. \qquad (13)$$

In order to determine the circle on the object plane (i.e., the circular feature), the following rotational transformation is applied to the current coordinate system [Reza Safaee-Rad et al. 1992]

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = T \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \qquad (14)$$

where $T = \begin{bmatrix} -\dfrac{m}{\sqrt{l^2+m^2}} & -\dfrac{nl}{\sqrt{l^2+m^2}} & l \\ \dfrac{l}{\sqrt{l^2+m^2}} & -\dfrac{mn}{\sqrt{l^2+m^2}} & m \\ 0 & \sqrt{l^2+m^2} & n \end{bmatrix}$.

The object plane (12) in the new coordinate $x'y'z'$ becomes

$$z' = nd_o \qquad (15)$$

This indicates that the object plane is perpendicular to the z' axis under the new coordinate system. Figure 2 shows the system after the transformation. Substituting the transformation (14) into the cone equation (11), we have

$$\begin{bmatrix} x' & y' & z' \end{bmatrix} \cdot Q_1 \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = 0 \qquad (16)$$

where $Q_1 = T' \cdot Q \cdot T$, and Q_1 can be written as

$$Q_1 = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \tag{17}$$

where,

$$q_{11} = a' f^2 \frac{m^2}{\tau^2} + b' f^2 \frac{l^2}{\tau^2} - 2h' f^2 \frac{ml}{\tau^2}$$

$$q_{12} = q_{21} = a' f^2 \frac{lmn}{\tau^2} - b' f^2 \frac{lmn}{\tau^2} + h' f^2 \frac{m^2 n}{\tau^2} - h' f^2 \frac{nl^2}{\tau^2} - g' fm + f' f$$

$$q_{13} = q_{31} = -a' f^2 \frac{ml}{\tau} + b' f^2 \frac{ml}{\tau} - h' f^2 \frac{m^2}{\tau} + h' f^2 \frac{l^2}{\tau} - g' f \frac{mn}{\tau} - f' f \frac{nl}{\tau}$$

$$q_{22} = a' f^2 \frac{l^2 n^2}{\tau^2} + b' f^2 \frac{m^2 n^2}{\tau^2} + d' \tau^2 + 2h' f^2 \frac{2mn^2 l}{\tau^2} - 2g' fnl - 2 f' fmn$$

$$q_{23} = q_{32} = -a' f^2 \frac{nl^2}{\tau} - b' f^2 \frac{m^2 n}{\tau} + d' \tau m - 2h' f^2 \frac{lmn}{\tau} - g' f \frac{n^2 l}{\tau} + g' fl\tau - f' f \frac{mn^2}{\tau} + f' fm\tau$$

$$q_{33} = a' f^2 l^2 + b' f^2 m^2 + d' n^2 + 2h' f^2 ml + 2g' fnl + 2 f' fmn$$

$$\tau = \sqrt{l^2 + m^2}$$

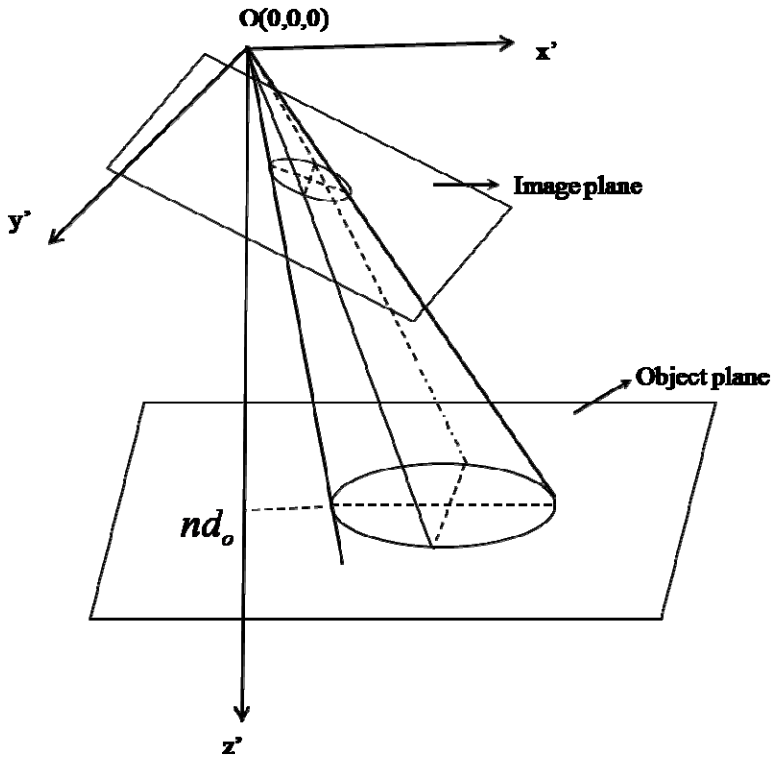

Figure 2.Diagram of the new coordinate system x'y'z' after rotation.

Now Eq. (16) represents the surface of the cone in the new coordinate system $x'y'z'$. From (15), we know that the object plane in this new coordinate system is vertical to the z' axis. If we substitute (15) to (16), the cone equation would reduce to the equation for the original circular reference feature. To be a circular equation, (17) must satisfy $q_{11} = q_{22}$ and $q_{12} = q_{21} = 0$, i.e.,

$$a' f^2 \frac{m^2}{\tau^2} + b' f^2 \frac{l^2}{\tau^2} - 2h' f^2 \frac{ml}{\tau^2} = a' f^2 \frac{l^2 n^2}{\tau^2} + b' f^2 \frac{m^2 n^2}{\tau^2} + d' \tau^2 + 2h' f^2 \frac{2mn^2 l}{\tau^2} - 2g' fnl - 2f' fmn$$

$$f\left(a' f \frac{lmn}{\tau^2} - b' f \frac{lmn}{\tau^2} + h' f \frac{m^2 n}{\tau^2} - h' f \frac{nl^2}{\tau^2} - g'm + f'l \right) = 0$$

If we define $k = \frac{f}{p_x}$, substituting (7) to above two equations, they can be reduced to,

$$t_1 k^2 \frac{m^2}{\tau^2} + t_3 k^2 \frac{l^2}{\tau^2} - 2t_2 k^2 \frac{ml}{\tau^2} = t_1 k^2 \frac{l^2 n^2}{\tau^2} + t_3 k^2 \frac{m^2 n^2}{\tau^2} + \tau^2 + 2t_2 k^2 \frac{mn^2 l}{\tau^2} - 2t_4 knl - 2t_5 kmn \quad (18)$$

$$(t_1 - t_3) k \frac{lmn}{\tau^2} + t_2 k \frac{m^2 n}{\tau^2} - t_2 k \frac{nl^2}{\tau^2} - t_4 m + t_5 l = 0 \quad (19)$$

In addition, we know from (17) and (13) that,

$$\tau = \sqrt{l^2 + m^2}$$

$$l^2 + m^2 + n^2 = 1$$

Analyzing above equations, we have four equations but five unknowns. Generally, there are infinite solutions. Solving them and presenting the explicit expressions of l, m, n may help us understand the property of the projection system, but will also be very difficult. However, I will leave this task for my future work because they do not contribute directly the current solutions. Here, I will first derive the relationship of l, m, n using the known vector in (6). Then, I will solve the problem by imposing certain constraints.

2.2 ORIENTATION ESTIMATION

From (18), we obtain,

$$n^4 - 2n^2 + 1 + 2(t_4 l + t_5 m)kn^3 + \left(t_1 k^2 l^2 + t_3 k^2 m^2 + 2t_2 k^2 ml\right)n^2 - 2(t_4 l + t_5 m)kn - t_1 k^2 m^2 - t_3 k^2 l^2 + 2t_2 k^2 ml = 0$$

which can be further reduced to

$$\left(n^2 - 1\right)^2 + 2\left(t_4 l + t_5 m\right)kn\left(n^2 - 1\right) + \left(t_1 k^2 l^2 + t_3 k^2 m^2 + 2t_2 k^2 ml\right)n^2 - t_1 k^2 m^2 - t_3 k^2 l^2 + 2t_2 k^2 ml = 0 \quad (20)$$

From (19), we obtain

$$n^2 - 1 = \frac{(t_1 - t_3)ml + t_2\left(m^2 - l^2\right)}{t_5 l - t_4 m}kn \quad (21)$$

.

Substituting (21) to (20), we obtain

$$\left(\frac{(t_1 - t_3)ml + t_2\left(m^2 - l^2\right)}{t_5 l - t_4 m}\right)^2 k^2 n^2 + \left(2\left(t_4 l + t_5 m\right)\frac{(t_1 - t_3)ml + t_2\left(m^2 - l^2\right)}{t_5 l - t_4 m} + t_1 l^2 + t_3 m^2 + 2t_2 ml\right)n^2 k^2$$
$$= \left(t_1 m^2 + t_3 l^2 - 2t_2 ml\right)k^2.$$

Eliminating k^2 on both sides and solving for n^2, we have,

$$n^2 = \frac{\left(t_1 m^2 + t_3 l^2 - 2t_2 ml\right)\left(t_5 l - t_4 m\right)^2}{b_1 l^4 + b_2 ml^3 + b_3 m^2 l^2 + b_4 m^3 l + b_5 m^4} \quad (22)$$

where

$$b_1 = t_1 t_5^2 - 2t_2 t_4 t_5 + t_2^2$$
$$b_2 = 2t_2 t_4^2 - 2t_3 t_4 t_5 - 2t_1 t_2 + 2t_2 t_3$$
$$b_3 = 2t_1 t_5^2 - t_3 t_5^2 - t_1 t_4^2 + 2t_3 t_4^2 + \left(t_1 - t_3\right)^2 - 2t_2^2$$
$$b_4 = 2t_2 t_5^2 - 2t_1 t_4 t_5 + 2t_1 t_2 - 2t_2 t_3$$
$$b_5 = t_3 t_4^2 - 2t_2 t_4 t_5 + t_2^2$$

Though we did not explicitly solve the equations, (22) can still help us obtain valuable solutions if we know any one value of l, m, n or a certain relationship between any two of them. Later, we will propose a way to solve for l, m, n in the general sense.

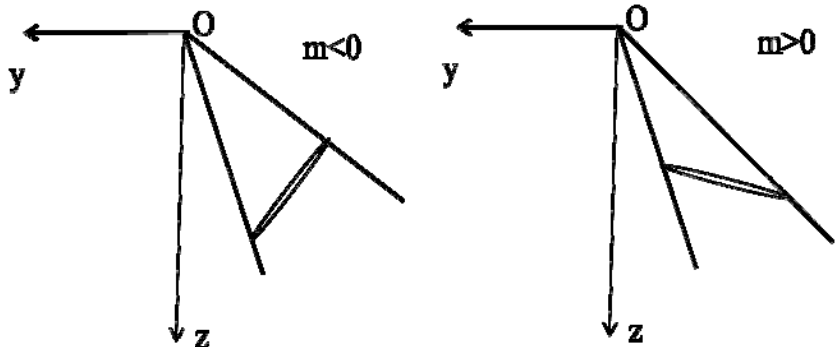

Figure 3. Projection of xyz along the x axis with two possible orientations.

2.2.1 One rotation angle

We consider the following two cases.

Case 1: $l = 0$

From (22), we obtain,

$$n = \sqrt{\frac{t_1 t_4^2}{t_3 t_4^2 - 2t_2 t_4 t_5 + t_2^2}} \qquad (23)$$

Here only the positive solution is retained because it only makes sense when the object is in front of the camera. Since (13) defines $l^2 + m^2 + n^2 = 1$, we can substitute (23) into (13) to solve for m. However, this would give us two solutions, one positive and the other negative. Figure 3 shows the schematic representation of the relative locations of the optical center and the

circular reference feature corresponding respectively to the two solutions. Whether only one solution should be kept is determined by the application. Because we mainly intend to apply our approach to the estimation of food portion sizes, only the positive solution is valid as shown in Fig. 3. Therefore, we obtain

$$m = \sqrt{1 - \frac{t_1 t_4^2}{t_3 t_4^2 - 2t_2 t_4 t_5 + t_2^2}} \tag{24}$$

In addition, from (19), we have

$$m = \frac{t_2}{t_4} kn \tag{25}$$

Substituting (25) and (23) into (13), we can solve for parameter k (the ratio of focal length and pixel size).

$$k = \sqrt{\frac{(t_3 - t_1)t_4^2 + t_2^2 - 2t_2 t_4 t_5}{t_1 t_2^2}} \tag{26}$$

Case 2: $m = 0$

Similarly,

$$n = \sqrt{\frac{t_3 t_5^2}{t_1 t_5^2 - 2t_2 t_4 t_5 + t_2^2}} \tag{27}$$

$$m = \sqrt{1 - \frac{t_3 t_5^2}{t_1 t_5^2 - 2t_2 t_4 t_5 + t_2^2}} \tag{28}$$

$$k = \sqrt{\frac{(t_1 - t_3)t_5^2 + t_2^2 - 2t_2 t_4 t_5}{t_2 t_3^2}} \tag{29}$$

Though we have discussed the above two cases separately, they are actually equivalent. When $l = 0$, the object plane is parallel to x axis; hence the semi-major axis is parallel to x axis. When $m = 0$, the object plane is parallel to y axis; hence the semi-major axis is parallel to y axis.

So for the ellipse equation (1), the coefficients of the x term when $l = 0$, are equivalent to those of y when $m = 0$. From (1) and (7), t_1 and t_4 correspond respectively to x^2 and x, and t_2, t_5 correspond to y^2 and y respectively. Therefore, the two cases are equivalent and the solution depends on how you define the x and y axes along the image plane.

Therefore, in this work, we consider the case of $l = 0$ and define the x and y axes in the direction shown in Figure 4.

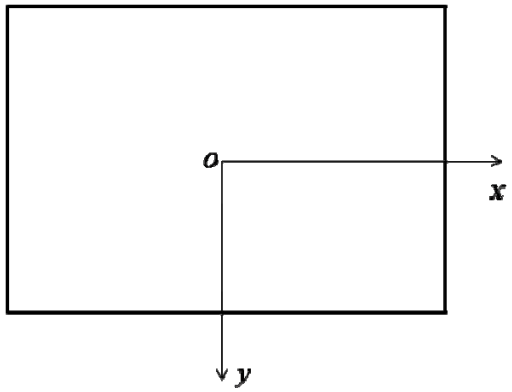

Figure 4. The directions of x and y axes mapped on the image.

The analytical solutions (24), (25) and (26) indicate that, given one image of a circular object whose surface is parallel to the x axis (Fig. 4), we are able to tell its orientation, and its focal length in terms of the pixel size.

2.2.2 Two rotation angles

What can be obtained for the case when the circular reference feature is placed in an arbitrary orientation? As shown in section 2.1, this case requires to solve for four variables l, m, n and k using three equations (13), (18) and (19). Though in general this case cannot be solved, appropriate solutions may exist if we further constrain that the optical axis of the camera passes

the center of the object plan, In this subsection, I would like to propose a method to approximate the solutions under this constraint which was inspired from the following experience: When we try to estimate the orientation of an object using one eye, we usually focus on the center of the object.

Figure 5 shows a typical perspective projection of a circular feature on the image when the optical axis passes the ellipse center and the object plane is not parallel to x axis. On this image, the angle of counterclockwise rotation from the x axis to the major axis of the ellipse is depicted by ϕ. The proposed method assumes that after the rotation of angle ϕ, in the new coordinate system $x_1 y_1 z_1$ (z is kept unchanged), the object plane is parallel to x_1. In other words, the major axis of the ellipse is parallel to the intersection line of the image and the object plane. Then, under the new coordinate system, the following derivation will be very similar to that in the previous case.

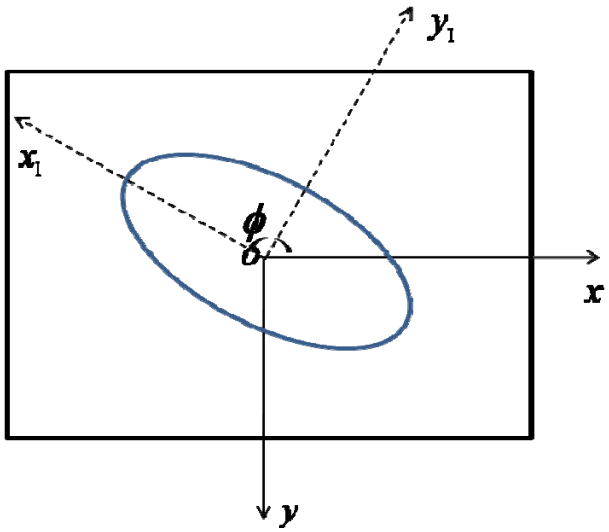

Figure 5. A perspective projection of a circular feature on the image.

For the ellipse of Eq. (1), the angle of ϕ can be determined by [Berger et al. 1984]

$$\phi = \begin{cases} 0 & h'=0, a'<b' \\ \dfrac{\pi}{2} & h'=0, a'>b' \\ \dfrac{1}{2}\cot^{-1}\dfrac{a'-b'}{2h'} & h'\neq 0, a'<b' \\ \dfrac{\pi}{2}+\dfrac{1}{2}\cot^{-1}\dfrac{a'-b'}{2h'} & h'\neq 0, a'>b' \end{cases} \qquad (30)$$

Combining with Eq. (7), Eq. (30) can be rewritten as a function of known values

$$\phi = \begin{cases} 0 & t_2=0, t_1<t_3 \\ \dfrac{\pi}{2} & t_2=0, t_1>t_3 \\ \dfrac{1}{2}\cot^{-1}\dfrac{t_1-t_3}{2t_2} & t_2\neq 0, t_1<t_3 \\ \dfrac{\pi}{2}+\dfrac{1}{2}\cot^{-1}\dfrac{t_1-t_3}{2t_2} & t_2\neq 0, t_1>t_3 \end{cases} \qquad (31)$$

To obtain the new coordinate system $x_1y_1z_1$, a rotation transformation is applied

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = T_1 \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}, \text{ where } T_1 = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ o & 0 & 1 \end{bmatrix} \qquad (32)$$

Substituting (32) into the cone equation (11), we obtain the cone equation under the new coordinate system

$$\begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix} \cdot Q_2 \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = 0 \qquad (33)$$

where,

$$Q_2 = T_1^{'} \cdot Q \cdot T_1 = \begin{bmatrix} q_{11}^{'} f^2 & q_{12}^{'} f^2 & q_{13}^{'} f \\ q_{21}^{'} f^2 & q_{22}^{'} f^2 & q_{23}^{'} f \\ q_{31}^{'} f & q_{32}^{'} f & q_{33}^{'} \end{bmatrix}$$

and

$$q_{11}^{'} = a'\cos^2\phi + b'\sin^2\phi + 2h'\sin\phi\cos\phi$$
$$q_{12}^{'} = q_{21}^{'} = (b'-a')\sin\phi\cos\phi + h'\cos 2\phi$$
$$q_{13}^{'} = q_{31}^{'} = g'\cos\phi + f'\sin\phi$$
$$q_{22}^{'} = a'\sin^2\phi + b'\cos^2\phi - 2h'\sin\phi\cos\phi$$
$$q_{23}^{'} = q_{32}^{'} = f'\cos\phi - g'\sin\phi$$
$$q_{33}^{'} = d'$$

Then, in this new coordinate system, if we define the orientation of the object plane as (l', m', n'), we apply the similar transformation on $x_1 y_1 z_1$ as equation (14)

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = T_2 \cdot \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}, \text{ where } T_2 = \begin{bmatrix} -\dfrac{m'}{\sqrt{l'^2+m'^2}} & -\dfrac{n'l'}{\sqrt{l'^2+m'^2}} & l' \\ \dfrac{l'}{\sqrt{l'^2+m'^2}} & -\dfrac{m'n'}{\sqrt{l'^2+m'^2}} & m' \\ 0 & \sqrt{l'^2+m'^2} & n' \end{bmatrix} \tag{34}$$

We have known that for the cone (11), the relationship of the three components of the object's orientation can be described as (22). Combining (7), Eq. (22) can be modified by

$$n^2 = \frac{(a'm^2 + b'l^2 - 2h'ml)(f'l - g'm)^2}{b_1 l^4 + b_2 ml^3 + b_3 m^2 l^2 + b_4 m^3 l + b_5 m^4} \tag{35}$$

where,

$$b_1 = a'f'^2 - 2h'g'f' + h'^2$$
$$b_2 = 2h'g'^2 - 2b'g'f' - 2a'h' + 2h'b'$$
$$b_3 = 2a'f'^2 - h'f'^2 - a'g'^2 + 2b'g'^2 + (a'-b')^2 - 2h'^2$$
$$b_4 = 2h'f'^2 - 2a'g'f' + 2a'h' - 2h'b'$$
$$b_5 = b'g'^2 - 2h'g'f' + h'^2$$

For the cone (33) in the new coordinate system $x_1 y_1 z_1$, the relationship of l', m', n' will be

$$n'^2 = \frac{(q_{11}^{'}m'^2 + q_{22}^{'}l'^2 - 2q_{12}^{'}m'l')(q_{23}^{'}l' - q_{13}^{'}m')^2}{b_1 l'^4 + b_2 m'l'^3 + b_3 m'^2 l'^2 + b_4 m'^3 l' + b_5 m'^4} \tag{36}$$

where,

$$b_1 = q_{11}^{'} q_{23}^{'\,2} - 2q_{12}^{'} q_{13}^{'} q_{23}^{'} + q_{12}^{'\,2}$$
$$b_2 = 2q_{12}^{'} q_{13}^{'\,2} - 2q_{22}^{'} q_{13}^{'} q_{23}^{'} - 2q_{11}^{'} q_{12}^{'} + 2q_{12}^{'} q_{22}^{'}$$
$$b_3 = 2q_{11}^{'} q_{23}^{'\,2} - q_{12}^{'} q_{23}^{'\,2} - q_{11}^{'} q_{13}^{'\,2} + 2q_{22}^{'} q_{13}^{'\,2} + \left(q_{11}^{'} - q_{22}^{'}\right)^2 - 2q_{12}^{'\,2}$$
$$b_4 = 2q_{12}^{'} q_{23}^{'\,2} - 2q_{11}^{'} q_{13}^{'} q_{23}^{'} + 2q_{11}^{'} q_{12}^{'} - 2q_{12}^{'} q_{22}^{'}$$
$$b_5 = q_{22}^{'} q_{13}^{'\,2} - 2q_{12}^{'} q_{13}^{'} q_{23}^{'} + q_{12}^{'\,2}$$

Substituting (7) and (33) into (36), we obtain

$$n'^2 = \frac{\left(c_1 m'^2 + c_4 l'^2 - 2c_2 m' l'\right)\left(c_5 l' - c_3 m'\right)^2}{\tau_1 l'^4 + \tau_2 m' l'^3 + \tau_3 m'^2 l'^2 + \tau_4 m'^3 l' + \tau_5 m'^4} \qquad (37)$$

where,

$$c_1 = t_1 \cos^2\phi + t_3 \sin^2\phi + 2t_2 \sin\phi\cos\phi$$
$$c_2 = \left(t_3 - t_1\right)\sin\phi\cos\phi + t_2 \cos 2\phi$$
$$c_3 = t_4 \cos\phi + t_5 \sin\phi$$
$$c_4 = t_1 \sin^2\phi + t_3 \cos^2\phi - 2t_2 \sin\phi\cos\phi$$
$$c_5 = t_5 \cos\phi - t_4 \sin\phi$$
$$\tau_1 = c_1 c_5^2 - 2c_2 c_3 c_5 + c_2^2$$
$$\tau_2 = 2c_2 c_3^2 - 2c_4 c_3 c_5 - 2c_1 c_2 + 2c_2 c_4$$
$$\tau_3 = 2c_1 c_5^2 - c_2 c_5^2 - c_1 c_3^2 + 2c_4 c_3^2 + \left(c_1 - c_4\right)^2 - 2c_2^2$$
$$\tau_4 = 2c_2 c_5^2 - 2c_1 c_3 c_5 + 2c_1 c_2 - 2c_2 c_4$$
$$\tau_5 = c_4 c_3^2 - 2c_2 c_3 c_5 + c_2^2$$

Which is independent from parameter $k = \frac{f}{p_x}$. Applying $l' = 0$ on (37) yields

$$n' = \sqrt{\frac{c_1 c_3^2}{\tau_5}}$$
$$m' = \sqrt{1 - \frac{c_1 c_3^2}{\tau_5}} \qquad (38)$$

Since the orientation $\left(0, \sqrt{1-\frac{c_1 c_3^2}{\tau_5}}, \sqrt{\frac{c_1 c_3^2}{\tau_5}}\right)$ is defined under the coordinate system $x_1 y_1 z_1$, we need to transform it back to the original coordinate system xyz. Applying (32), we have

$$\begin{bmatrix} l \\ m \\ n \end{bmatrix} = \left(-\sin\phi \cdot \sqrt{1-\frac{c_1 c_3^2}{\tau_5}}, \cos\phi \cdot \sqrt{1-\frac{c_1 c_3^2}{\tau_5}}, \sqrt{\frac{c_1 c_3^2}{\tau_5}} \right)^T \qquad (39)$$

Substituting (39) back into (19), we obtain

$$k = \frac{\left(t_4 m - t_5 l\right)\left(l^2 + m^2\right)}{\left(t_1 - t_3\right) lmn + t_2 n\left(m^2 - l^2\right)} \qquad (40)$$

Therefore, when the ellipse center is located at the center of the image, the orientation of the circular reference feature and the parameter k can also be calculated.

2.3 OBJECT DIMENSION ESTIMATION

From above derivation, we have obtained analytical expressions of l, m, n and k. Our goal is to perform dimensional measurements on the object plane. Obviously, these parameters are insufficient to determine the location of the object plane because we would get the same parameters by moving the object plane back and forth in parallel. Therefore, we still need some additional information to determine the object's location in order to conduct the desired dimensional measurements. In this work, we assume that the size of the physical circular feature is given. Now we derive closed-form solutions for the estimation of dimensions on the object plane.

In order to use the information of the size of the reference circular feature, we need to obtain the standard form of the circle equation. As discussed above, (15) indicates that the object plane in the new coordinate system $x'y'z'$ is perpendicular to the z' axis. Therefore, if we substitute (15) into cone equation (17), we would obtain the equation for the circular reference feature. Combining with (7), the circle equation can be written as follows

$$e_1 x'^2 + e_1 y'^2 - 2e_2 d_0 x' + 2e_3 d_0 y' + e_4 d_0^2 = 0 \qquad (41)$$

where

$$e_1 = k^2\left(t_1 m^2 + t_3 l^2 - 2t_2 ml\right)\left(m^2 + l^2\right)^{-1}$$
$$e_2 = kn\left(m^2 + l^2\right)^{-1/2}\left[kml\left(t_1 - t_3\right) + kt_2\left(m^2 - l^2\right) + t_4 mn - t_5 nl\right]$$
$$e_3 = kn^2\left(m^2 + l^2\right)^{-1/2}\left(-t_1 kl^2 - t_3 km^2 - 2t_2 kml - t_4 nl - t_5 mn\right) + n\left(m^2 + l^2\right)^{1/2}\left(n + t_4 kl + t_5 km\right)$$
$$e_4 = \left(t_1 k^2 l^2 + t_3 k^2 m^2 + n^2 + 2t_2 k^2 ml + 2t_4 knl + 2t_5 kmn\right)n^2$$

Eq. (41) can be further expressed in the standard form of a circular function

$$\left(x' - \frac{e_2}{e_1} d_0\right)^2 + \left(y' + \frac{e_3}{e_1} d_0\right)^2 = \left(\frac{e_2^2 + e_3^2}{e_1^2} - \frac{e_4}{e_1}\right) d_0^2 \qquad (42)$$

If we know that the diameter, D, of the circular feature, we then have

$$d_0 = \frac{D}{2} \cdot \left(\frac{e_2^2 + e_3^2}{e_1^2} - \frac{e_4}{e_1}\right)^{-1/2}. \qquad (43)$$

The center of the physical circular feature in the coordinate system $x'y'z'$ is

$$\left(x_0', y_0', z_0'\right)^T = \left(\frac{e_2}{e_1} d_0, -\frac{e_3}{e_1} d_0, n d_0\right)^T \qquad (44)$$

Substituting (44) into (14), we obtain its coordinates under the original coordinate system

$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} e_1^{-1}\left(m^2 + l^2\right)^{-1/2}\left(nle_3 - me_2\right) + nl \\ e_1^{-1}\left(m^2 + l^2\right)^{-1/2}\left(mne_3 - le_2\right) + mn \\ -e_1^{-1}\left(m^2 + l^2\right)^{1/2} + n^2 \end{bmatrix} \qquad (45)$$

Since d_0 is known, the distance of the object plane to the origin point in $x'y'z'$, nd_0 will be known. Therefore, the location of the object plane is determined.

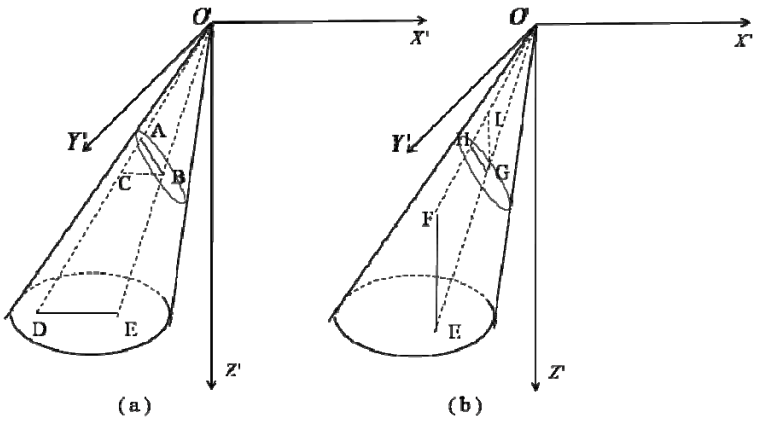

Figure 6. (a) dimension of the object plane; (b) dimension perpendicular to the object plane

2.3.1 Dimension on reference plane

Comparing Figures 1 and 2, since the object plane in the new coordinate system $x'y'z'$ as shown in Figure 2 is perpendicular to the optical axis, intuitively it appears to be easy to establish a geometric model in the transformed coordinate system $x'y'z'$. Let us first consider the dimension on the object plane (Fig. 6(a)), we assume that points A and B in the original coordinate system are $A:(x_1, y_1, f)$ and $B:(x_2, y_2, f)$. The supplemental line BC is parallel to the desired dimension DE, where point C is on line O'AD. We now determine the dimensional length $|DE|$.

The coordinates of points A, B in the transformed coordinate system are:

$A:(-x_1, -ny_1 + mf, my_1 + nf)^T$ and $B:(-x_2, -ny_2 + mf, my_2 + nf)^T$.

The equation of line $O'D$ is given by

$$\frac{x'}{-x_1} = \frac{y'}{-ny_1 + mf} = \frac{z'}{my_1 + nf} \qquad (46)$$

Since $BC \,//\, DE$, the point C has the same z' coordinate as the point B

$$z_c' = my_2 + nf \qquad (47)$$

Substituting (47) into (46), we obtain

$$x_c' = -x_1 \cdot \frac{my_2 + nf}{my_1 + nf}$$

$$y_c' = (-ny_1 + mf) \cdot \frac{my_2 + nf}{my_1 + nf}$$

Then,

$$|BC| = \sqrt{\left(x_2 - x_1 \cdot \frac{my_2 + nf}{my_1 + nf}\right)^2 + \left(-ny_2 + mf - (-ny_1 + mf) \cdot \frac{my_2 + nf}{my_1 + nf}\right)^2} \qquad (48)$$

Also, because,

$$\frac{|BC|}{|DE|} = \frac{my_2 + nf}{nd_0} \qquad (49)$$

we can solve for $|DE|$

$$|DE| = |BC| \cdot nd_0 \cdot (my_2 + nf)^{-1} \qquad (50)$$

Substituting (2) into (50), we have

$$|DE| = nd_0 \cdot (m(y_{p2} - c_2) + nk)^{-1} \cdot \sqrt{\left((x_{p2} - c_1) - (x_{p1} - c_1) \cdot \frac{m(y_{p2} - c_2) + nk}{m(y_{p1} - c_2) + nk}\right)^2 + \left(-n(y_{p2} - c_2) + mk - (-n(y_{p1} - c_2) + mk) \cdot \frac{m(y_{p2} - c_2) + nk}{m(y_{p1} - c_2) + nk}\right)^2} \qquad (51)$$

2.3.2 Vertical dimension to reference plane

As shown in (51), we have the closed-form solution for the estimation of the dimension on the object plane. Then for dimensions parallel to the object plane but not on it, can we still estimate them? Obviously, this is just the case with a constant vertical distance (along the z'-axis in the $x'y'z'$ coordinate system) shifted by the object plane. This vertical distance, which often corresponds to the height estimation of an object, is important in calculating the volume of the object, as in the case of food portion size estimation.

Let us consider the diagram in Figure 6(b). Line HG is the perspective projection of the vertical dimension EF on the image plane. The supplemental line GI is parallel to EF, and the point I is on line $O'F$. In the original coordinate system, we define $G:(x_3, y_3, f)$ and

$H:(x_4, y_4, f)$. Then, in $x'y'z'$ their coordinates are $G:(-x_3, -ny_3 + mf, my_3 + nf)^T$ and $H:(-x_4, -ny_4 + mf, my_4 + nf)^T$ .The equation of line $O'F$ is given by

$$\frac{x'}{-x_4} = \frac{y'}{-ny_4 + mf} = \frac{z'}{my_4 + nf} \qquad (52)$$

In the coordinate system $x'y'z'$, the object plane is perpendicular to z'axis. Then, the vertical dimension FE is parallel to the z' axis. Hence, the supplemental line GI is perpendicular to the object plane and parallel to the z' axis, and point I has the same x' and y' coordinates of point G . Substituting the x' and y' coordinates back into (52), we notice there are two equalities that can be used to solve for the z' coordinate:

$$\frac{-x_3}{-x_4} = \frac{z_I'}{my_4 + nf} \qquad (53)$$

$$\frac{-ny_3 + mf}{-ny_4 + mf} = \frac{z_I'}{my_4 + nf} \qquad (54)$$

Theoretically, (53) and (54) should result in the same z_I'. However, numerically we may obtain different values. Because essentially we want to use the coordinate difference of the point I and G to calculate z_I', we should use the equation providing the higher difference between the numerator and denominator of its left side. In our application to food volume estimation, the plate of food is usually on a table and the vertical dimension of a food object tends to align along the y axis of the image (Fig. 4). Then, we use (54) to calculate z_I'.

$$z_I' = (my_4 + nf) \cdot \frac{-ny_3 + mf}{-ny_4 + mf} \qquad (55)$$

Then

$$|GI| = (my_3 + nf) - (my_4 + nf) \cdot \frac{-ny_3 + mf}{-ny_4 + mf} \qquad (56)$$

From $\frac{|GI|}{|EF|} = \frac{z_G^{'}}{z_E^{'}}$, we obtain

$$|EF| = nd_0 \cdot \left(\frac{my_4 + nf}{my_3 + nf} \cdot \frac{-ny_3 + mf}{-ny_4 + mf} - 1 \right) \qquad (57)$$

Substituting (2) into (57) yields

$$|EF| = nd_0 \cdot \left(1 - \frac{m(y_{p4} - c_2) + nk}{m(y_{p3} - c_2) + nk} \cdot \frac{-n(y_{p3} - c_2) + mk}{-n(y_{p4} - c_2) + mk} \right) \qquad (58)$$

Therefore, we can also calculate the dimension which is vertical to the object plane. Equivalently, we are able to calculate dimensions perpendicular to the object plane, such as the height of a regularly shaped object.

3.0 EXPERIMENT AND RESULT ANALYSIS

We have shown that the orientation of the object plane can be estimated if a single image is given containing a circular feature with its plane parallel to x axis. This estimation does not need to know the focal length and pixel size of the camera. Additionally, if the optical axis of the camera passes through the center of the circular feature, the orientation of the object plane can also be calculated. If we are given the size of the physical circular feature, we are able to calculate dimensions of the object. In this section, we conduct both synthetic and real experiment to test and verify the above derivation. Figure 7 shows a typical image of 2-D patterns used in our experiments, where several circular reference features in different quadrants are plotted. The circular shapes in different locations were used in order to comprehensively verify our approach and analyze the effect of the location of reference feature on the estimation result. In the experiment, the center of the pattern was placed on the optical axis in order to better control the rotation angle. Figure 8 shows the equipment used in our experiment. We will test and analyze the case of rotation in one direction first, and then the case of flexible orientation of the object plane.

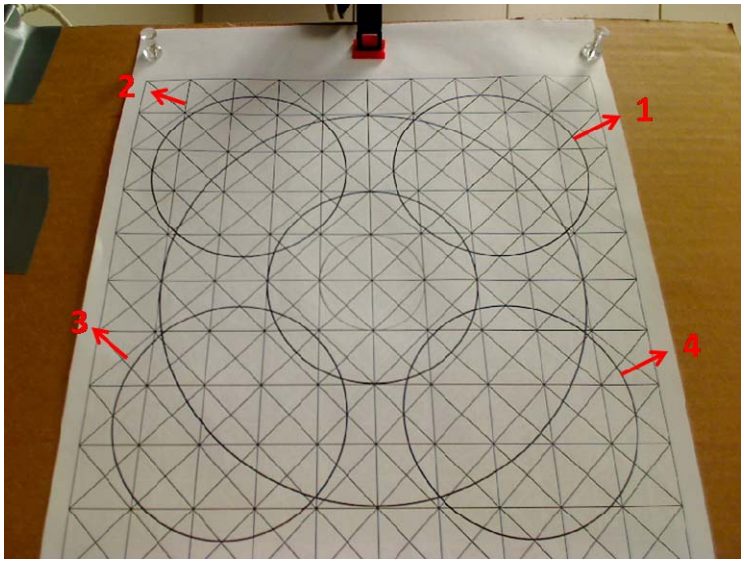

Figure 7. Typical image of planar patterns used in experiment.

Figure 8. To be changed for the equipment.

3.1 ONE ROTATION ANGLE

We have set up two separate experiments to capture images of the patterns printed on a piece of paper at different rotation angles and the distances relative to the camera conducted by two different persons at different days. Within both experiments, we respectively used the four circles, No. 1-4 in fig. 7, in the four quadrants as the reference. For each calculation, we first selected eight points on the ellipse boundary and used the least square method to fit an ellipse. After we obtain the ellipse equation in the pixel coordinates, five points which evenly distributed on the ellipse were randomly generated as the input to our approach. As shown in Figure 9, when we used the circular reference feature in certain quadrant, we calculated the solid bold line in the corresponding quadrant for the first experiment and the dashed bold diameter for the second experiment. Figure 10-13 presents the calculated results of the rotation angle, dimension, distance and the ratio of the focal length to the pixel size. Note that the rotation angle can be easily calculated from the orientation vector. For each image, our method was applied using the circular shape in each of the four quadrants and gave four results as shown in the figures. Table 1 shows the mean error of calculation.

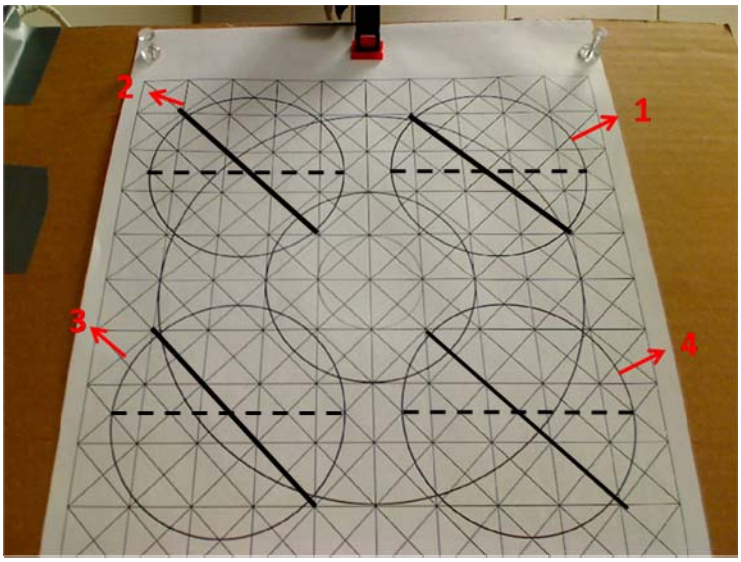

Figure 9. Linear lengths measurement.

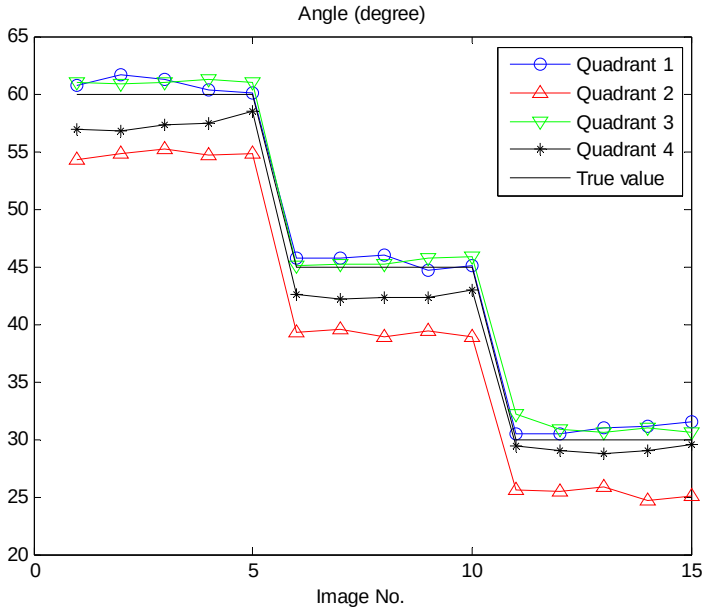

Figure 10. First experiment: results of estimated rotation angle.

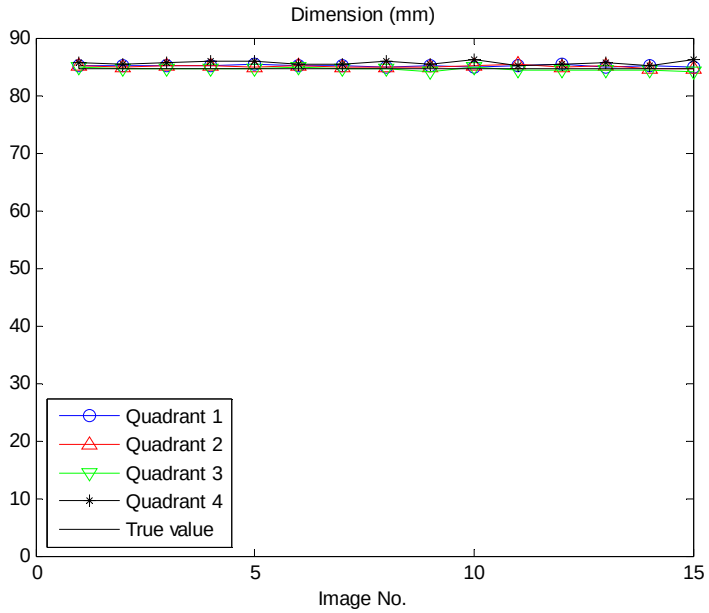

Figure 11. First experiment: results of estimated linear length.

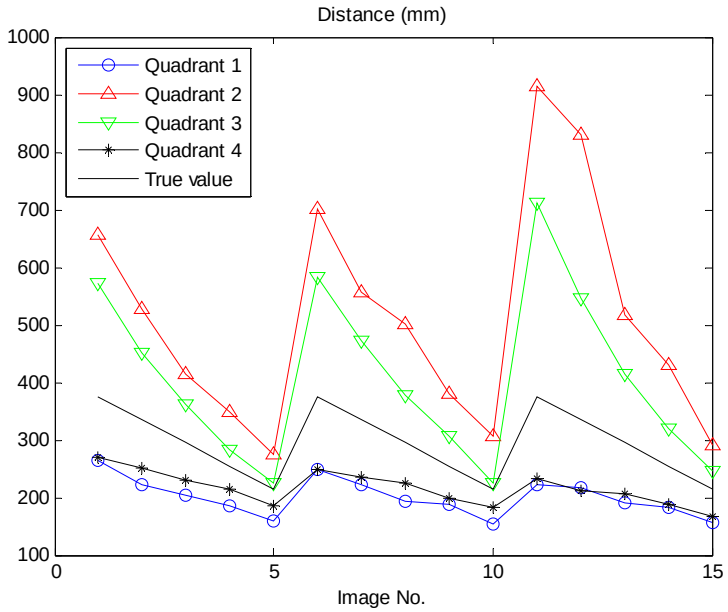

Figure 12. First experiment: results of estimated distance.

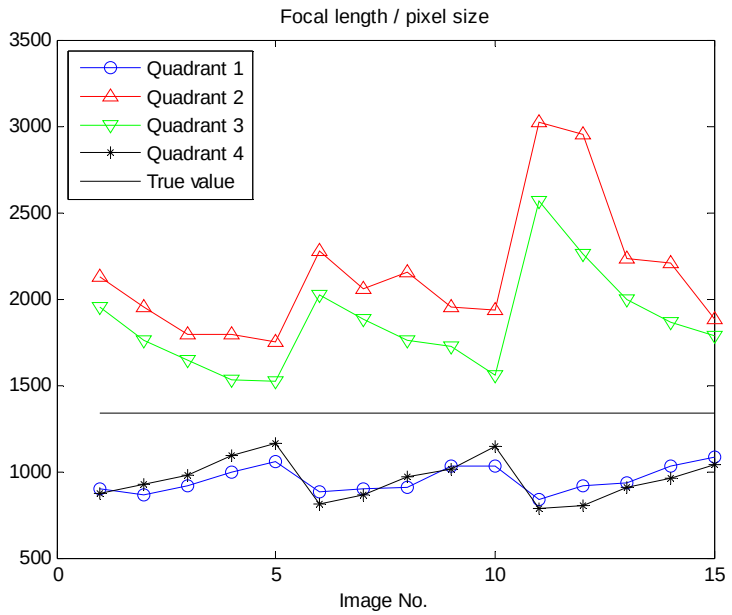

Figure 13. First experiment: results of estimated ratio of the focal length to the pixel size.

Table 1. Mean error using the circular reference in each quadrant for the first experiment

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Angle	1.77%	-12.41%	2.04%	-4.30%
k	-28.60%	60.42%	39.08%	-28.35%
Distance	-31.29%	68.31%	33.94%	-25.42%
Dimension	0.70%	0.57%	0.00%	1.29%

Similarly, for the second experiment, Figure 14-17 presents the results of estimated rotation angle, dimension, distance and ratio of focal length to pixel size. Table 2 lists the mean error with respect to each variable.

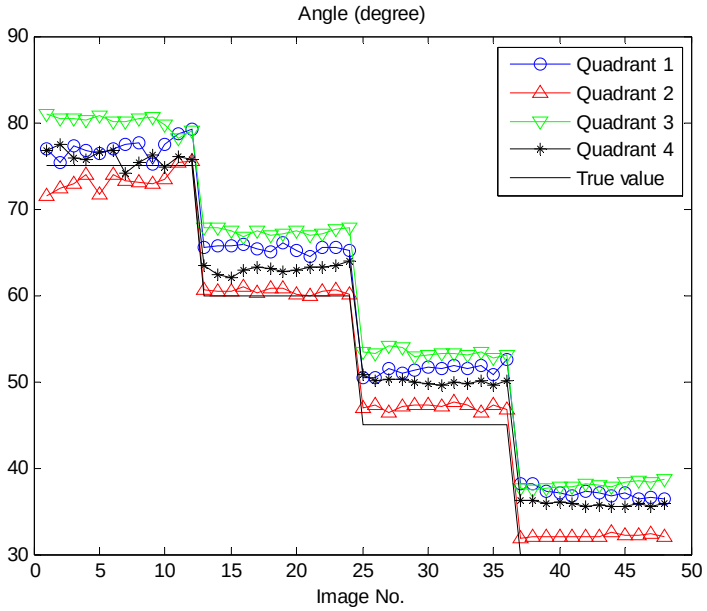

Figure 14. Second experiment: results of estimated rotation angle.

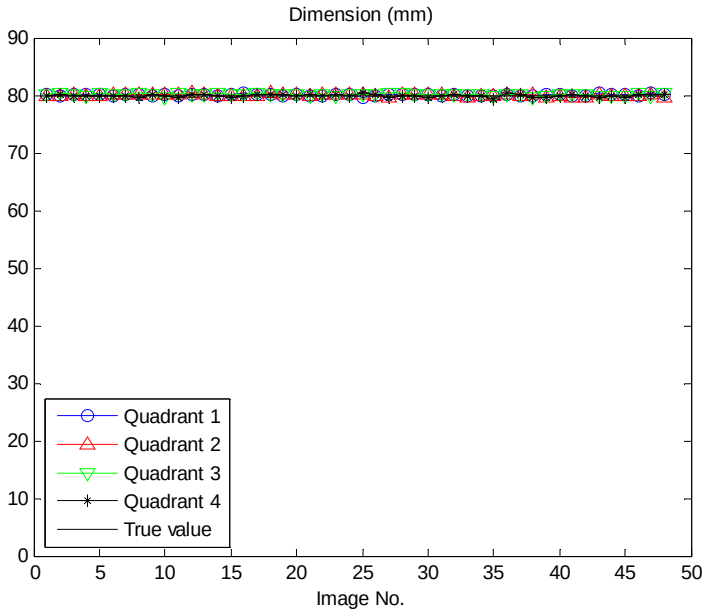

Figure 15. Second experiment: results of estimated linear length.

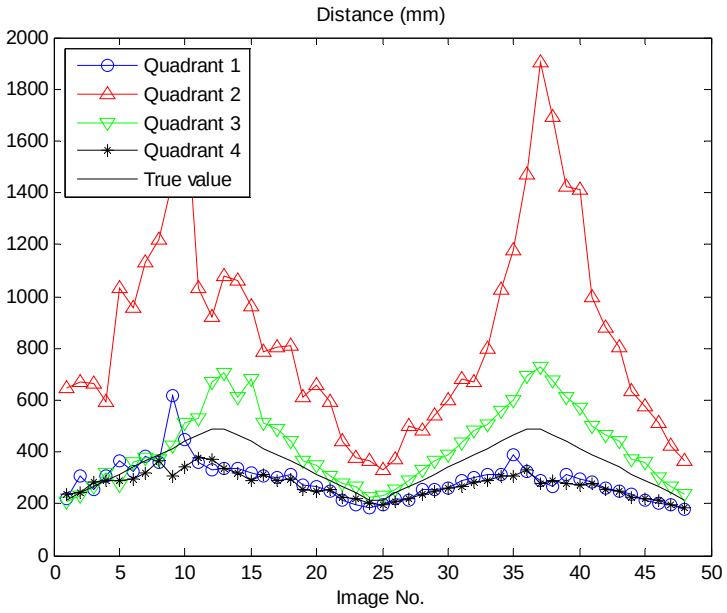

Figure 16. Second experiment: results of estimated distance.

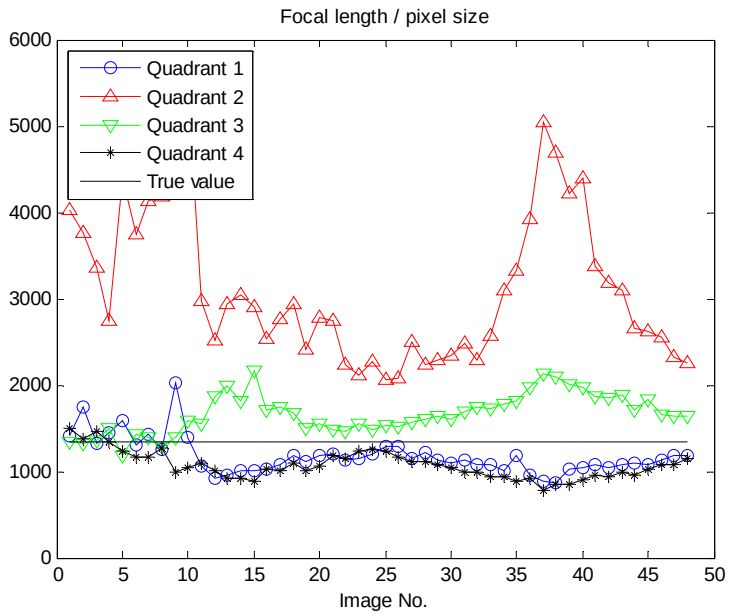

Figure 17. Second experiment: results of estimated ratio of the focal length to the pixel size.

Table 2. Mean error using the circular reference in each quadrant for the second experiment

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Angle	12.42%	2.52%	16.00%	9.23%
k	-11.92%	133.24%	24.76%	-19.75%
Distance	-31.29%	68.31%	33.94%	-25.42%
Dimension	0.03%	-0.05%	0.14%	-0.08%

By comparing the results of the first and second experiments shown in Figs. 10-13 and Figs 14-17, respectively, considerable differences in different quadrants can be observed except for the case of dimensional estimation. Large errors are also observed in Tables 1 and 2 with respect to the estimations of the rotation angle, distance, and ratio between focal length and pixel size. Despite the relatively large errors in these variables, the final result of dimensional estimation is highly accurate. In order to understand the sources of the large errors, we conducted a sensitivity analysis on the selection of the eight points to fit the ellipse, i.e., the sensitivity to ellipse equation in pixel coordinates. We chose one image, and for the ellipse in each quadrant, a

series of noise with normal distribution N(0,3) was added to the pixel coordinates of selected points. Figures 18-21 show the results for each quadrant after 500 noise perturbations. In the chart of each variable, the middle black horizontal line represents the mean value and the upper and lower red horizontal lines represent the mean plus and minus the standard deviation, respectively. Table 3 provides the ratio of the standard deviation to the mean. This ratio indicates the estimation sensitivity to the perturbation.

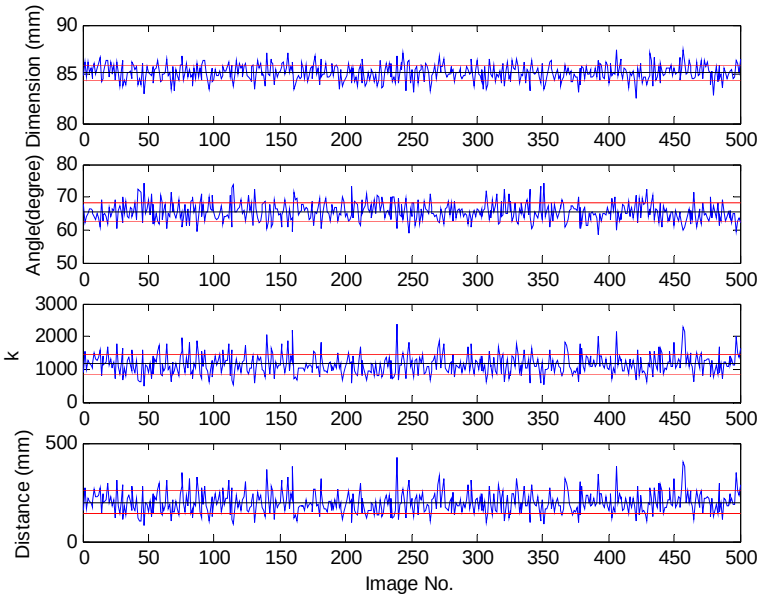

Figure 18. Sensitivity analysis using the circular reference in quadrant 1.

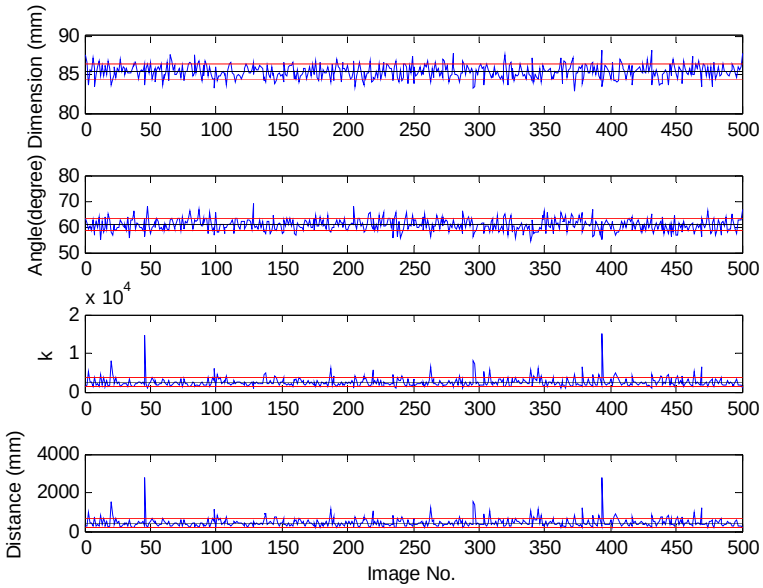

Figure 19. Sensitivity analysis using the circular reference in quadrant 2.

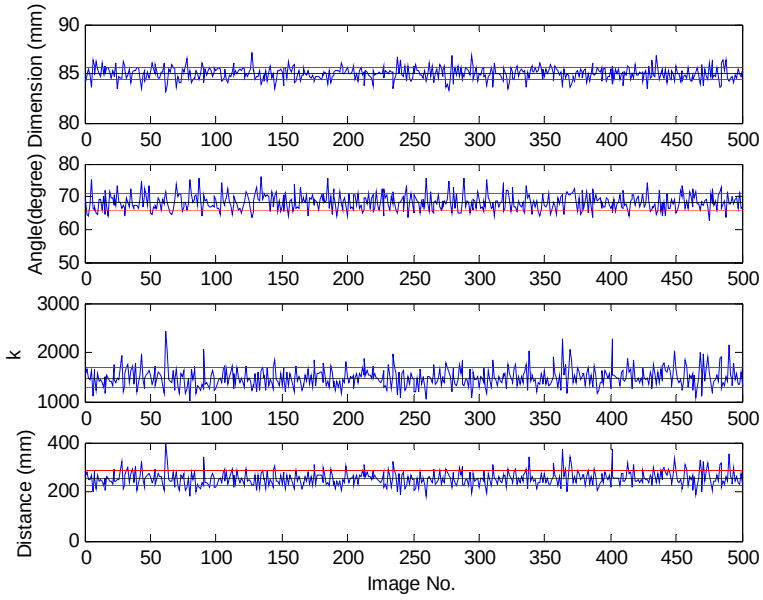

Figure 20. Sensitivity analysis using the circular reference in quadrant 3.

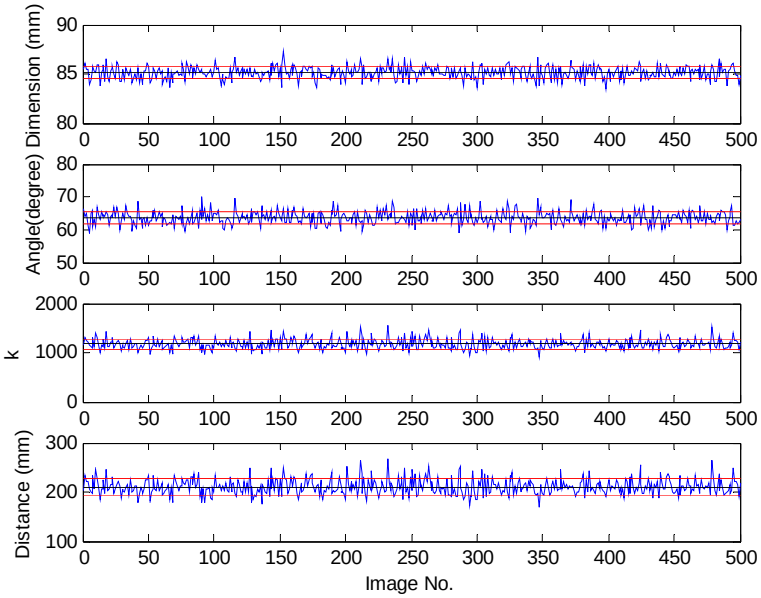

Figure 21. Sensitivity analysis using the circular reference in quadrant 4.

Table 3. The ratio of standard deviation and mean.

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Dimension	0.98%	1.06%	0.75%	0.74%
Angle	4.35%	3.94%	3.62%	3.08%
Ratio k	26.97%	51.50%	13.33%	8.77%
Distance	28.41%	53.93%	11.97%	7.55%

From Figures 10-21 and Tables 2-3, it can be observed that the dimensional measurement is accurate and stable to the perturbation of the ellipse equation. However the intermediate parameters, including the ratio of focal length to pixel size, distance, and rotation angle are inaccurate. Among the three intermediate parameters, the error in the rotation angle, usually within ten percent, is relatively acceptable. However, the results for the other two parameters are beyond our acceptance and they are also very sensitive to the noise. Although the above analysis shows that noise can cause significant error and instability, it still cannot account for the inconsistent results of using circular references in different quadrants. The error of our

equipment (an instrumentation lathe) and experimental setup were also sources of error because we could hardly control preciously the object board. Despite these error sources, we found that the idealized pin-hole model was the primary source of error because it did take the distortion into account, resulting in poor performances in the estimation of the intermediate parameters. To eliminate the effect of equipment setup, distortion and the manual selection of points to fit ellipse, we performed a simulation study. Images were generated by the ideal pinhole model without distortion compensation from the physical locations of the pattern board and the camera in our second experiment. Figure 22 shows a typical simulated pattern imposed on the original image. Figures 23-26 show the simulation results for all four parameters. Table 5 shows the mean errors for the simulation results. Figures 23-26 and Table 5 indicate that our approach is accurate under the ideal pin-hole model condition. In other words, the large error in the results of the three intermediate parameters, ratio of focal length and pixel size, distance and rotation angle were caused by the idealized pin-hole model. Though distortion can be compensated when using pinhole model as discussed in [Safaee-Rad 1992], we must pre-calibrate the camera's intrinsic parameters. However, in many applications, we do not know these camera parameters and avoiding using these parameters is our interest. Therefore, a different approach must be taken to solve this problem. We will leave the study of this problem as a future work since, in the present study, we are interested only in dimensional estimation. Our results have shown that the estimation results with respect to this particular variable are highly accurate.

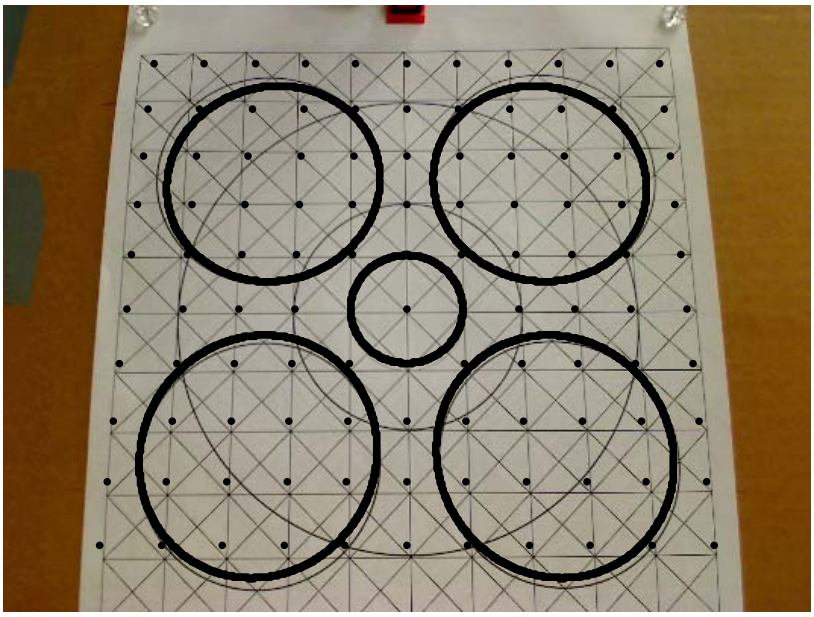

Figure 22. Simulated pattern imposed on the original image.

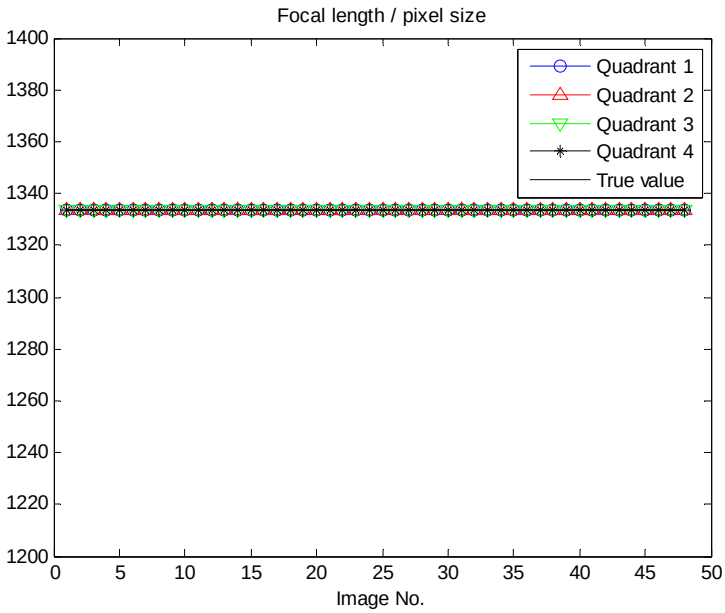

Figure 23. Simulation results of k on 48 images.

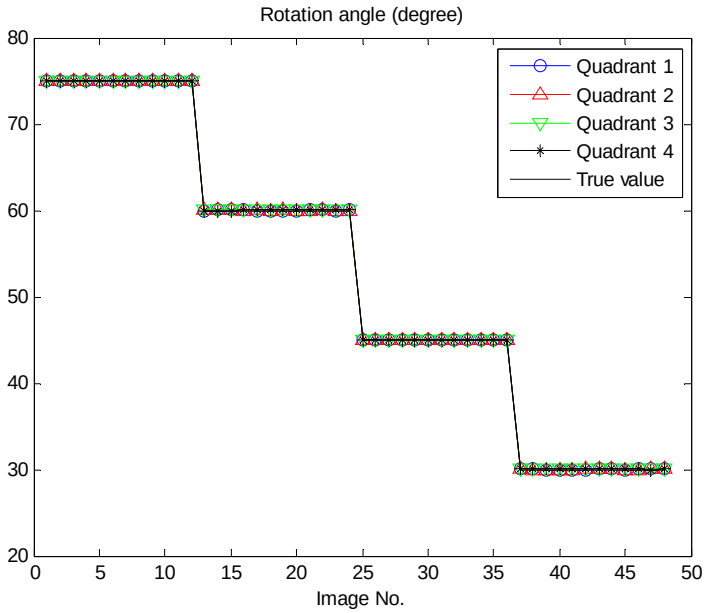

Figure 24. Simulation results of the rotation angle on 48 images.

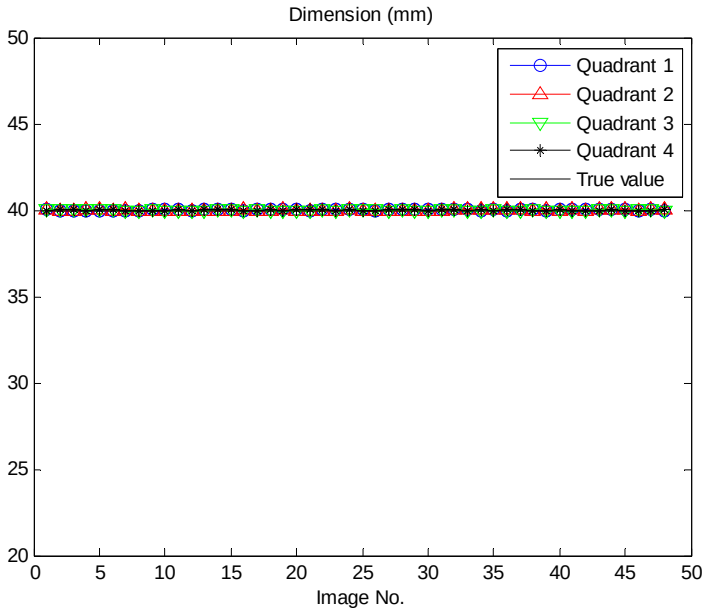

Figure 25. Simulation results of the linear length on 48 images.

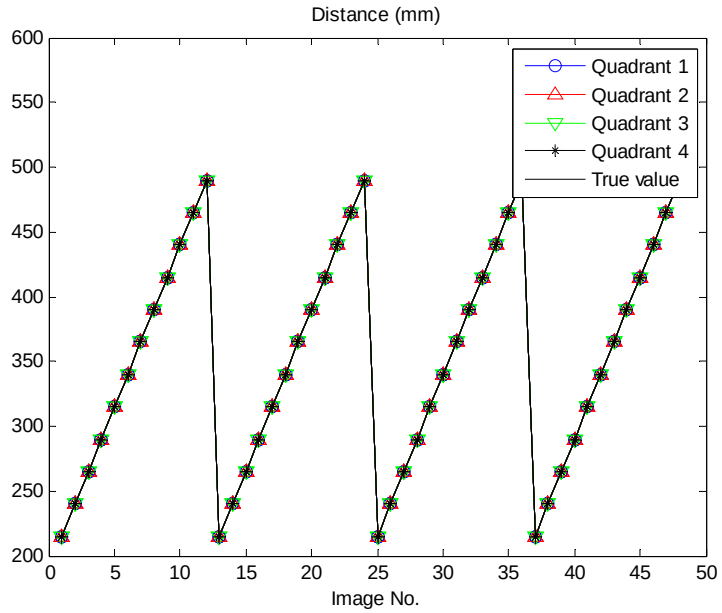

Figure 26. Simulation results of distance on 48 images.

Table 4. Mean error using the circular reference in each quadrant for the synthetic experiment.

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Angle	0%	0%	0%	0%
k	0%	0%	0%	0%
Distance	0%	0%	0%	0%
Dimension	0%	0%	0%	0%

3.2 TWO ROTATION ANGLES

In order to evaluate our method for the case where the circular reference plane is allowed to take an arbitrary orientation, we adjusted our pattern paper board and let it have two different rotational angles relative to the camera. Images were captured from varying distances. We used the central circular shape as reference in the calculation as shown in Figure 9. Figure 27 shows the results of two rotation angles. The upper panel in Figure 27 represents the rotation angle around the x-axis while the lower panel corresponds to the rotation around the y-axis. Since the experiment with two rotations is more difficult to control than that with one rotation, the estimation errors were larger. A simulation study was therefore conducted to analyze the source of error. Figure 28 shows the calculation result based on the simulated images. From Figure 27-28s, we find that indeed the rotation angles can be approximated by using the deformation of the ellipse only when the optical axis passes through the center of the circular reference. Also, we notice that, as the pattern board rotates around the y-axis, the accuracy of estimation involving two rotation angles decreases.

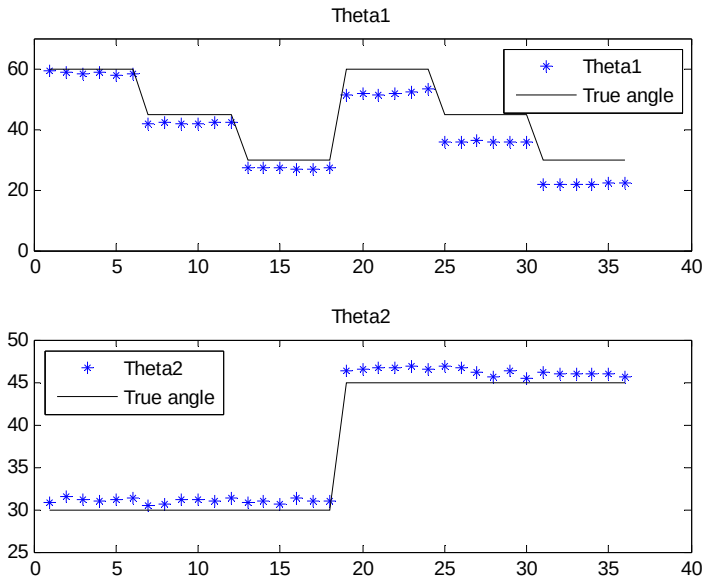

Figure 27. Results of the two rotation angles using the central circular reference.

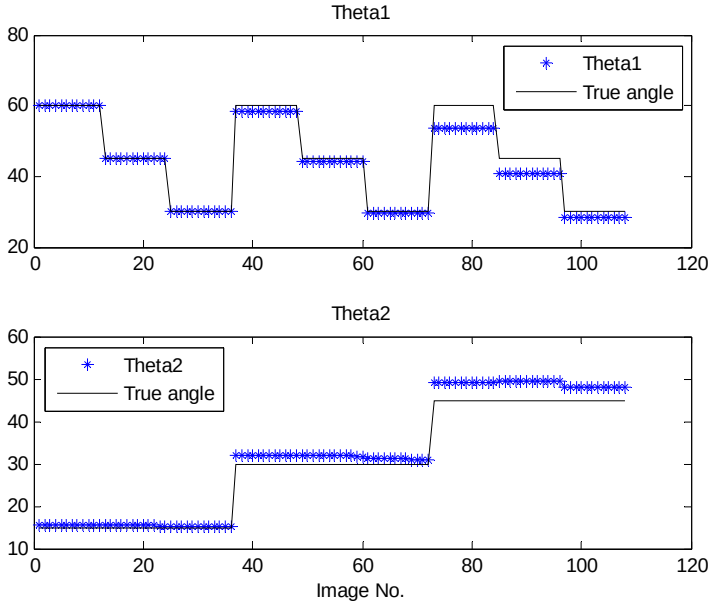

Figure 28. Simulation results of the two rotation angles.

4.0 APPLICATION

The estimation of food portion size is very important in the study of obesity. The image of a food or an eating event usually involves one or more circular features, such as a dining plate, a bowl or a cup. The size of a selected circular feature can be measured by the human subject before or after the dietary study. Therefore, using a circular feature as a reference allows convenient dietary study when a wearable camera is used to take pictures automatically at a certain rate. This type of study is objective and passive to the human subject since he or she does not need to place a reference on the dining table.

Figures 29-31 show the pictures of different foods. These pictures were taken using two digital cameras, both having the auto-focus function and picture resolution of 2048*1536. When images were acquired, we rotated the camera only in one direction. Figures 32-34 show the results of estimation. Table 5 provides more detailed performance information.

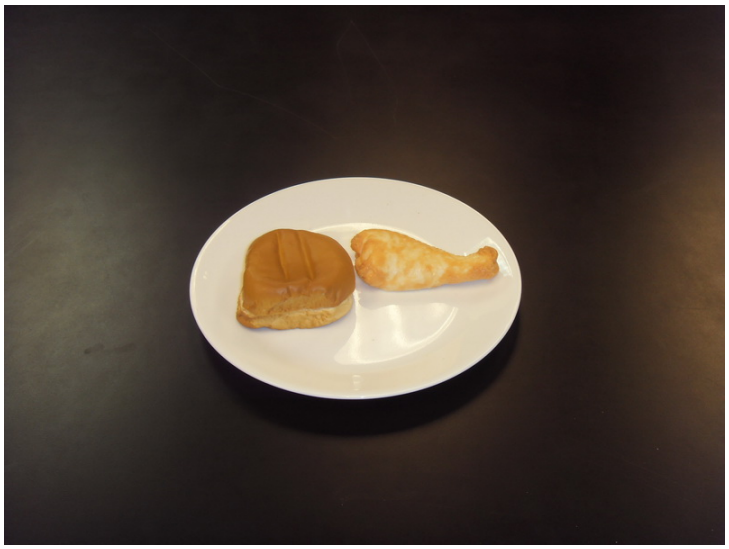

Figure 29. Roll and chicken leg used in the estimation.

Figure 30. Bread and cake used in the estimation.

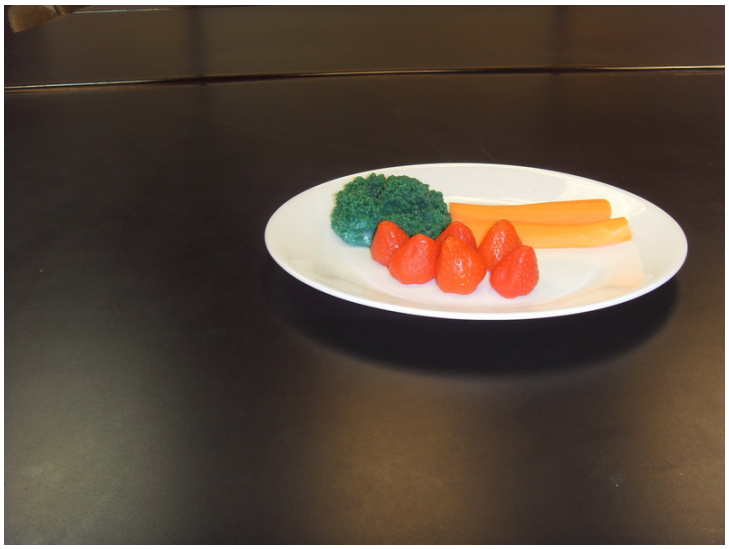

Figure 31. Carrot and strawberry used in the estimation.

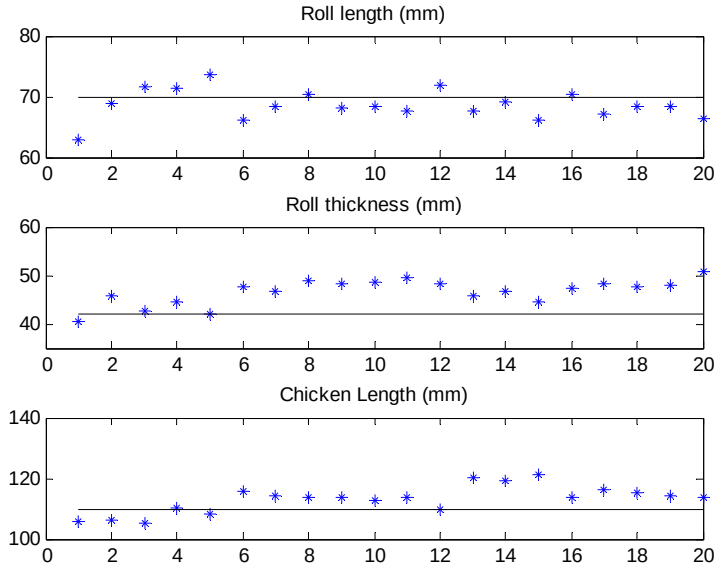

Figure 32. Results correspond to foods in Fig. 29. Line indicates the ground truth.

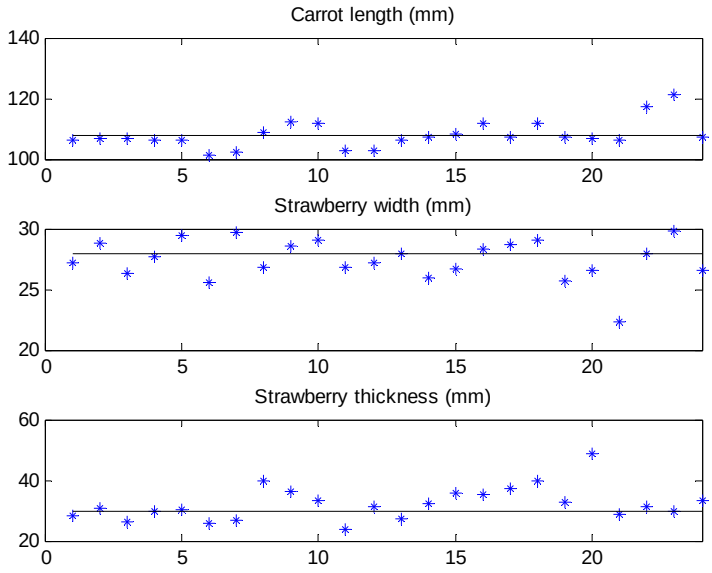

Figure 33. Results correspond to foods in Fig. 30. Line indicates the ground truth.

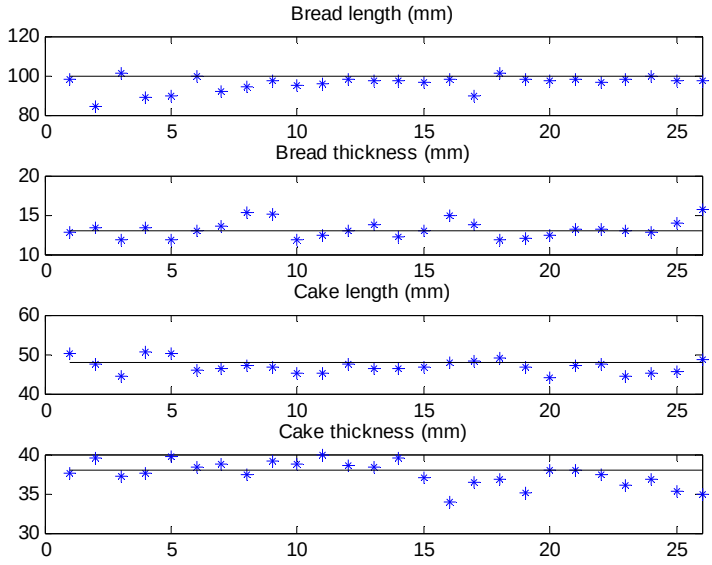

Figure 34. Results correspond to foods in Fig. 31. Line indicates the ground truth.

Table 5. Estimation results of food dimensions.

Food	Dim. Type	Sample Size	Truth (mm)	Mean (mm)	Std.(mm)	Error (%)
Roll	Length	20	70	68.63	2.45	-1.95
	Thickness	20	42	46.65	2.63	11.08
Strawberry	Length	24	28	27.44	1.69	-2.01
	Thickness	24	30	32.22	5.55	7.38
Cake	Length	26	48	46.98	1.79	-2.13
	Thickness	26	38	37.56	1.58	-1.16
Bread	Length	26	100	96.03	4.09	-3.97
	Thickness	26	13	13.19	1.07	1.49
Chicken	Length	20	110	113.17	4.46	2.88
Carrot	Length	24	108	108.05	4.50	0.05

From Figures 32-34 and table 5, we can see that the estimation results of the food dimensions are very good. The thicknesses of the strawberry and the roll have relative lower accuracy. However, if we revisit Figures 29-31, it can be easily observed that the vertical dimensions for these two types of foods are relatively harder to measure. The selection of the starting and ending points of the vertical dimension for these foods was ambiguous. Even with this difficulty, the results were still satisfactory for the case of food volume estimation.

5.0 DISCUSSION

Our analysis and experiments have shown that the proposed approach provides us with satisfactory results in dimensional estimation using a circular feature and a single image input. Though the closed-form solutions are mainly derived under the constraint that there is only one relative rotation angle between the camera and the circular reference plane, the approach still has many applications. We point out that the assumption of one rotation angle between the camera and the circular reference is realistic. For example, when a person adjusts the location and zoom of his/her camera in or out to optimizing the scene in the picture, he/she usually only rotates the camera towards or outwards the body. A parallel motion of the camera will not affect the application of our method as long as there is only one relative rotation angle between the camera and the circular feature. In robotic applications, the camera can be controlled more precisely to satisfy the constraint. In automatic inspection of the manufacturing process, this condition can also be satisfied easily. Our method also works when the second rotation angle is small. In the food portion size estimation study, the pictures are captured by a hand-held digital camera. Therefore, some small rotation along the other orthogonal direction is expected. However, as indicated by our experimental results, the error caused by camera rotation in the undesired direction was small, and the accuracy of our food portion size estimates was sufficient. Despite these observations, the sensitivity and tolerance of undesirable camera rotation in the dimensional estimation is worthy of further study in the future.

In our experiments, we have found that there are two special cases where our approach cannot handle well when the object plane rotates in one direction. These cases are:

1) When the reference plane is vertical to the optical axis, the perspective projection of a circular reference feature is still a circle. In this case, no information from the shape deformation can be used to estimation the location of the reference plane.

2) When the center of the circular feature is located along the y-axis, the ellipse is deformed only along the direction of y-axis. Then we lose one dimensional information, and hence cannot calculate the orientation of the reference plane as well as the parameters that require the orientation.

We point out that these cases correspond to $t_4 = 0$ in (23). Hence we have $n = 0$ and consequently Eq. (25) can no longer be used to calculate k.

When there are two relative rotation angles between the camera and the circular reference feature, certain special cases need to be discussed. As shown in the previous section, when the optical axis roughly passes through the center of the circular feature, we can approximately estimate the orientation of the reference plane. In practical cases, if the camera can be controlled, its location or/and orientation can be adjusted to satisfy that the optical axis passes through the center of the reference feature. Then, the two relative rotation angles can be approximately calculated. The camera can be further adjusted to create the condition that there is only one rotation angle between the camera and the circular reference feature. However, when the center of the perspective projection of the circular reference, that is the ellipse, deviates considerably from the center of the image, the approach cannot estimate its rotation angles. This is because the deformation of the ellipse is highly related to its location in the image which cannot be covered only by the rotation of the coordinate system in the proposed method.

Finally, our approach has provided closed-form solutions for the three intermediate parameters, including the orientation of the circular reference feature, its distance to the camera and the ratio of the focal length to the pixel size. However, our experiments have shown that, because of the effect of image distortion, these three parameters cannot be calculated accurately. However, the orientation of the reference plane can be estimated with an error usually below 10%. However, an error up to 20% may be observed occasionally.

6.0 CONCLUSION

In this thesis, a fundamental computer vision problem is revisited to estimate the dimensions of objects in one single image without knowing the focal length and the pixel size of the camera. A new approach is proposed based on a simple pin-hole camera model and a circular reference feature in the image to estimate the dimensions of objects in the same image. Under the constraint that only one relative rotation angle is allowed between the camera and the circular reference feature, a set of closed-form solutions was derived for the ratio between the focal length and the pixel size, the orientation of the reference feature plane, the distance from the reference feature to the camera, and the object dimensions. Our method is highly accurate in estimating object dimensions, but less accurate in estimating the ratio between the focal length and the pixel size, the orientation of the reference feature plane, and the distance from the reference feature to the camera. We have found that the estimation error is mainly due to the distortion of the pinhole model. Our method has been applied to the estimation of food portion size using a dining plate as a reference feature. We have estimated volumes of different types of foods, including chicken leg, carrot, dinner roll, strawberry, cake and bread. Our estimation results are satisfactory, providing a useful measurement tool for the study of obesity. Although this work can be used to solve many practical problems, further study is necessary to improve estimation accuracy and extend the method to less restrictive cases.

BIBLIOGRAPHY

M.A. Fishler and R.C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. Ass. Compu. Mach*., vol. 24, no 6, pp. 381-395, June 1981.

M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis, and Machine Vision, Thomson Engineering*, 2007.

R. Horaud, B. Conio, and O. Leboulleux, “An analytic solution for the perspective 4-point problem,” *Comput. Vision, Graphics Image Processing*, vol. 47, pp. 33-44, 1989.

M. Kabuka and A.E. Arenas, “Position verification of a mobile robot using standard pattern,” *IEEE J. Robotics and Automation*, vol. 3, no. 6, pp. 505-516, Dec. 1987.

R.M. Haralick, “Using perspective transformations in scene analysis,” *Comp Vision, Graphics, and Image Processing*, vol. 13, no. 3, pp191-221, 1980.

R.M. Haralick, “ Solving camera parameters from the perspective projection of a parameterized curve,” *Pattern Recognit*., vol. 17, no. 6, pp. 637-645, April. 1984.

R.M. Haralick, “Determining camera parameters from the perspective projection of a rectangle,” *Pattern Recognition*, vol. 22, no. 3, pp. 225-230, 1989.

S. Linnainmaa, D. Harwood, and L.S. Davis, “Pose determination of a three-dimensional object using triangle pairs,” *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, pp. 634-647, 1988.

W. J. Wolfe, D. Mathis, C. W. Sklair, and M. Magee, “The perspective view of three points,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 66-73, Jan. 1991.

W. J. Wolfe and K. Jones, “Camera calibration using the perspective view of a triangle,” in *Proc. SPIE Conf. Automated Inspection Measurement*, vol. 730, Cambridge, MA, Oct. 28-30, 1986.

D. DeMenthon and L. S. Davis, “Exact and approximate solutions of the perspective-three-point problem,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. no. 11, pp 1100-1105 Nov. 1992.

M.A. Abidi and T. Chandra, “A new efficient and direct solution for pose estimation using quadrangular targets: algorithm and evaluation.” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 534-538, May. 1995.

X.S. Gao, X. R. Hou, J. Tang and H.F. Cheng, “ Complete solution classification for the perspective-three-point problem”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930-943, Aug. 2003.

M. J. Magee and J. K. Aggarwal, “Determining the position of a robot using a single calibration object,” in *Prof. IEEE Int. Conf. Robotics Automat.* (Atlanta, GA), pp. 140-149, Mar. 1984.

Shanti Narayan, *Analytical Solid Geometry*, 12th ed. S.Chand And Company, 1961.

Y. C. Shin and S. Ahmad, “3D location of circular and spherical features by monocular model-based vision,” in *Proc. IEEE Int. Conf. Syst. Man Cybern.* (Boston), Nov. 1989, pp. 576-581.

R. Safaee-Rad, I. Tchoukanov, K. C. Smith, and B. Benhabib, “Three-Dimensional location estimation of circular features for machine vision,” *IEEE Trans. Robotics Automat.*, vol. 8, no. 5, pp. 624-640. Oct. 1992.

P.G. Mulgaonkar, “Analysis of perspective line drawings using hypothesis based reasoning,” Ph.D dissertation, Virginia Polytechnic Institute and State University, 1984.

D. H. Marimont, “Inferring spatial structure from feature correspondences,” Ph.D dissertation, Stanford University, Stanford, CA, March 1986.

H. S. Sawhney, J. Oliensis, and A. R. Hanson, “Description from image trajectories of rotational motion,” in *Prof. 3rd IEEE Int. Conf. Comput. Vision* (Osaka, Japan), pp. 494-498, Dec. 1990.

J. Trabulsi, D. A. Schoeller, “Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake,” *Am J Physiol.* Vol. 281, pp.E891-E899., 2001

J. Ngo, A. Engelen, M. Molag, J. Roesle, P. García-Segovia and L. Serra-Majem, “*A review of the use of information and communication technologies* for dietary assessment,” *Br J Nutr.*, vol. 101, pp. S102-S112, 2009

M. Sun, J. Fernstrom, W. Jia, S.A. Hackworth, N. Yao, Y. Li, C. Li, M. Fernstrom, R. J. Sclabassi, “A wearable electronic system for objective dietary assessment,” *J Am Diet Assoc.*, vol. 110, pp. 45-47, 2010.

A. Fitzgibbon, M. Pilu, and R. B. Fisher, “Direct least square fitting of ellipses,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476-480, May. 1999.

M. Berger, P. Pansu, J. P. Berry, X.Saint-Raymond, and S. Levy, “Problems in geometry”, New York: Springer-Verlag, 1984.