

FOOD DIMENSION ESTIMATION FROM A SINGLE IMAGE USING STRUCTURED LIGHTS

by

Ning Yao

Master in Electrical Engineering, Hong Kong University of Science
and Technology, 2005

Submitted to the Graduate Faculty of
the Swanson School of Engineering in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH
SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Ning Yao

It was defended on

October 18, 2010

and approved by

Mingui Sun, Department of Neurological Surgery, Bioengineering and Electrical
Engineering, University of Pittsburgh

Robert J. Sclabassi, Department of Neurological Surgery, Bioengineering and Electrical
Engineering, University of Pittsburgh

Ching-Chung Li, Department of Electrical and Computer Engineering, University of
Pittsburgh

Luis F. Chaparro , Department of Electrical and Computer Engineering, University of
Pittsburgh

Zhi-Hong Mao, Department of Electrical and Computer Engineering, University of
Pittsburgh

Dissertation Director: Mingui Sun, Department of Neurological Surgery, Bioengineering
and Electrical Engineering, University of Pittsburgh

FOOD DIMENSION ESTIMATION FROM A SINGLE IMAGE USING STRUCTURED LIGHTS

Ning Yao, PhD

University of Pittsburgh, 2010

Two-thirds the population in the United States of America are overweight or obese. The annual medical expenditures attributable to obesity may be as high as \$215 billion per year. Obesity has been linked to many types of diseases, including cancer, type 2 diabetes, cardiovascular diseases, stroke and birth defects. Deaths related to obesity are estimated at 300,000 each year in the United States. In order to understand the etiology of the obesity epidemic and develop effective weight management methods for obese patients, accurate dietary data is an essential requirement. However, the current dietary assessment methods, which depend on self-reported data by the respondents, have an estimated 20% to 50% discrepancy from the intake. This large error severely affects obesity research.

The recent rapid advances in electrical engineering and information technology fields have provided sophisticated devices and intelligent algorithms for dietary assessment. Considering portability and battery-life, systems installed with a single camera have the advantages of low cost, space saving, and low power consumption. Although several methods have been proposed to estimate food quantities and dimensions, many of these methods cannot be used in practice because of their inconvenience, and the requirement of calibration and maintenance. In this dissertation, we present several approaches to food dimensional estimation using two types of structured lights. These approaches are low in cost and power consumption, and suitable for small and portable image acquisition devices.

Our first design uses structured laser beams as reference lights. Three identical laser modules are structured to form an equilateral triangle on the plane orthogonal to the camera

optical axis. A new method based on orthogonal linear regression is proposed to release restrictions on the laser beams, so that the precision requirement for equilateral triangle can be relaxed. Based on the perspective projection geometry, intersections of structured laser beams and perspective projection rays are estimated, which construct a spatial plane containing the projection of the objects of interest. The dimensions of the objects on the observed plane are then calculated. In the second design, an LED diode is used as a reference light. A new algorithm is developed to estimate the object plane using the deformation of the observed ellipse.

In order to provide a precise system calibration between the structured lights and the camera, an orthogonal linear regression method is proposed to calibrate the structured lights. Characteristics of the reference features are investigated. A color-based thresholding method is proposed to segment features. An ellipse fitting method is used to extract feature parameters. The extraction results of our algorithms are very close to those manually performed by human.

Several experiments are performed to test our designs using both artificial and real food. Our experimental results show an average estimation error of less than 10%.

TABLE OF CONTENTS

1.0	INTRODUCTION	1
1.1	Significance of The Work	1
1.2	Existing Image-based Approaches to Food Dimensional Estimation	3
1.3	Other Potential Applications of Proposed Dimension Estimation Techniques	5
1.4	Contribution of This Dissertation	6
2.0	BACKGROUND	8
2.1	Basic Concepts of the Pinhole Camera model	8
2.1.1	Perspective and Affine Spaces	9
2.1.2	Camera Intrinsic and Extrinsic Parameters	11
2.1.2.1	Intrinsic Parameters	11
2.1.2.2	Extrinsic Parameters	13
2.2	Review of 3D Reconstruction Methods	14
2.2.1	3D Reconstruction from Range Scanner	14
2.2.2	3D Reconstruction from Stereo or Successive Image Sequences	15
2.2.3	Structured Lights Aided 3D Reconstruction	16
2.2.4	3D Reconstruction from a Single Image	17
2.3	Alternative Methods for 3D Reconstruction from a Single Image	18
2.3.1	Shade from Shading and Texture	18
2.3.2	Vanishing Point	19
2.3.3	Statistical Models	20
2.3.4	Focusing and Defocusing Methods	20
2.3.5	Structured Light Methods	21

2.3.6	Circular Feature Methods	21
3.0	MEASUREMENT SYSTEM DESIGN	23
3.1	Design of Wearable Device for Dietary Assessment	23
3.2	Laser Beam Based Structured Light Design	24
3.3	LED Based Structured Light Design	26
3.4	System Integration	26
4.0	GEOMETRIC DIMENSIONAL ESTIMATION ALGORITHMS BASED ON PERSPECTIVE AND STRUCTURED LIGHTS	28
4.1	Assumptions and Geometric Relations among Image, Camera Plane, Laser Beams and Object Plane	28
4.2	Ideal Laser Beam Model and Corresponding Algorithm	29
4.3	Modified Laser Beam Model and Corresponding Algorithm	33
4.3.1	Estimation of Approximate Intersections' Coordinates and Planar Patch Equation	33
4.4	Assumptions and Geometric Relation among Image, Camera Plane, LED Beam and Object Plane	35
4.5	LED Beam System Model and Corresponding Algorithm	36
4.6	Height and Volume Estimation of Regularly Shaped Objects	37
5.0	SYSTEM CALIBRATION	40
5.1	Camera System Calibration	40
5.1.1	Checkerboard Approach	40
5.2	Structured Lights Calibration	43
5.2.1	Laser Beams Calibration	43
5.2.1.1	Orthogonal Linear Regression Method	43
5.2.1.2	Measurement Of Spatial Line Equation of $O'A'$, $O'B'$ and $O'C'$	44
5.2.2	LED Beam Calibration	46
6.0	CIRCULAR PATTERN EXTRACTION	48
6.1	Techniques for Circular Feature Pattern Extraction	48
6.1.1	Pattern Segmentation	48
6.1.2	Feature Vector Extraction	49

6.2	Color-based Thresholding Segmentation and Pattern Extraction	51
6.2.1	Laser Pattern Extraction	51
6.2.2	LED Pattern Extraction	54
7.0	EXPERIMENTAL RESULTS	63
7.1	Evaluation of Laser Beams Design	63
7.1.1	Experiment with High Resolution Digital Camera	63
7.1.2	Experiment with Miniature Camera	68
7.2	LED Beam Design Experiment	73
7.3	Discussion	76
8.0	CONCLUSION	78
9.0	FUTURE WORK	81
9.1	More Efficient System Design and Feature Extraction Difficulties	81
9.2	Robust Dimension Estimation	82
10.0	PUBLICATIONS	83
	APPENDIX. GRADIENT DESCENT SEARCHING METHOD	85
	BIBLIOGRAPHY	88

LIST OF TABLES

1	Distance and physical length estimation with artificial objects	67
2	Real food physical length estimation at arbitrary position using laser structured light.	67
3	Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 20° with respect.	69
4	Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 20° with respect (cont.). . . .	70
5	Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 50° with respect.	70
6	Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 50° with respect (cont.). . . .	71
7	Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 90° with respect.	71
8	Estimated distances from perpendicular object plane	73
9	Estimated distances and angles from tilted object planes	73
10	Dimension and height estimates using LED structured light at different distances and a plane tilted by 20°	74
11	Dimension and height estimates using LED structured light at different distances and a plane tilted by 50°	75
12	Searching the approximated intersection point between two skew lines	86
13	Searching the approximated intersection point between two skew lines (Continued)	87

LIST OF FIGURES

1	Statistics of overweight in the United States (Source from <i>CDC National Center for Health Statistics</i>)	2
2	Data flow of the entire system in [27]	4
3	Process of the food portion estimation in [19]	4
4	(a) Rays of light travel from the object through the picture plane to the viewer's eye O . (b)Space principle of a pinhole camera. (Courtesy of [47])	9
5	Frontal pinhole imaging model	10
6	Image coordinate frame origin translation.	12
7	Stereoscopic camera and stereoscopic viewer	15
8	Optical triangulation. (a) 2D triangulation using a laser beam for illumination; (b) Extension to 3D; (c) Red laser line projected onto small (20cm) statuette; (d) Reflected light captured by CCD camera. (Courtesy of [60])	17
9	3D reconstruction of a car seat (Photo courtesy of [47])	18
10	From a monocular view with a constant intensity light source (left figure) and a single distant light source of known incident orientation upon an object with known reflectance map (right figure)	19
11	Prototype of wearable electronic device	23
12	Laser point-array pattern generated by a diffraction grating glass	24
13	Experimental instruments with three laser diodes	26
14	Structured lights with object of interest	27
15	Geometric relationships among camera plane, object plane, and camera origin	29
16	Geometric relationship of ideal laser beams design	30

17	Six skew lines and estimate intersections	34
18	Geometrical relationships among image, camera plane, LED spotlight and object plane	35
19	Height estimation on a perpendicular plane S_2	38
20	Intrinsic and extrinsic parameter calibrations	42
21	Structured light calibration system setup	45
22	Sampling along the z direction	46
23	Spatial lines fitting using Orthogonal Linear Regression	47
24	LED diode calibration experimental setup, measurement and fitting	47
25	Intensity comparison of channel splitting : (a) A laser spot on a white paper background in Red, Green, and Blue channels. (b)A laser spot on a white paper background in Hue, Saturation, and Value channels.	53
26	3D shapes of intensities in Red, Green, and Blue channels.	54
27	Performance comparison on wheat bread: (a) Histogram-based method; (b)Clustering-based method.	55
28	Performance comparison on yellow cheese: (a) Histogram-based method; (b)Clustering-based method.	56
29	Performance comparison on a plate of noodle and beans: (a) Histogram-based method; (b)Clustering-based method.	57
30	Segmentation on Red channel, Difference-Image, and Value channel and automatic ellipse-fitting are shown on the Red channel and Value channel.	58
31	Intensity comparison of channel splitting : (a) An LED spot on a white paper background in Red, Green, and Blue channels. (b)An LED spot on a white paper background in Hue, Saturation, and Value channels.	59
32	Performance comparison: (a) Histogram-based method; (b) Clustering-based method.	60
33	Performance comparison: (a) Histogram-based method; (b) Clustering-based method.	61
34	Segmentation and automatic ellipse fitting on Red, Green, Hue and Saturation channels.	62

35	Prototype with laser beams structured light and high resolution digital camera used in our experiment.	64
36	Distance and dimension estimation on a perpendicular plane	65
37	Real food samples for the dimension estimation experiments.	66
38	Prototype with a miniature camera.	68
39	Real objects used for dimension and height estimation	72
40	Flowchart of the proposed food dimension estimation approach	78

1.0 INTRODUCTION

1.1 SIGNIFICANCE OF THE WORK

More than 190 million Americans, or two-thirds of the population, are overweight or obese, which is doubled approximately in the last thirty year. Moreover, more than 17% of children in the United States, about 12.5 million, are overweight or obese, which has tripled in the last thirty years [1] (Figure 1). Annual medical expenditures attributable to obesity have doubled in less than a decade, and may be as high as \$147 billion per year, according to a new study by researchers at RTI International, the Agency for Healthcare Research and Quality, and the U.S. Centers for Disease Control & Prevention, published in the health policy journal *Health Affairs*. Obesity has been linked to many types of cancers (e.g., breast, colon, and prostate cancers) [2], type 2 diabetes [3], cardiovascular diseases [4], stroke [6], digestive diseases [5], respiratory diseases [7], osteoarthritis [8], and birth defects [10]. In the U.S., obesity related deaths are estimated to be 300,000 each year [9].

One simple but critical cause of being overweight or obese is when a person's intake energy exceeds the amount of energy used. A lack of physical activity is also an important cause [11]. In order to understand the etiology of the obesity epidemic in the U.S. and to develop effective weight management methods for obese patients, accurate acquisition of diet data from free-living individuals is an essential requirement. Current dietary assessments are largely dependent on self-reported data by respondents. Standard 24-hour recalls itemize all food and nutrients consumed during the previous reporting day, and food frequency questionnaires detail the intake frequency of a long list of foods over a specific time [12]. Several dietary assessment tools, such as CalorieKing [13] and CalorieCounter [14], can provide simple calorie calculation based on food portion size. However, they are of limited use

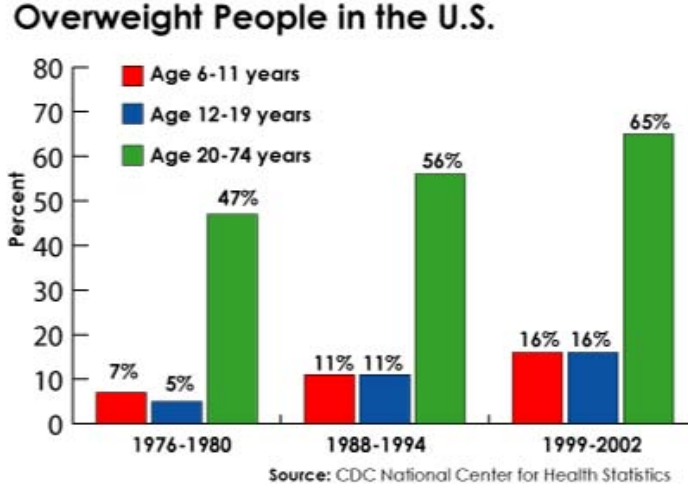


Figure 1: Statistics of overweight in the United States (Source from *CDC National Center for Health Statistics*)

since it is difficult for users to accurately estimate food sizes. Other forms of dietary recalls and questionnaires have also been designed to meet specific research needs [16]. Although these methods have been commonly adopted in various research and public health settings, nutritional scientists have questioned whether the self-reported data truly reflects the amount of energy the respondents habitually ingest due to a significant under-reporting [15]. It has been shown that the discrepancy of energy between the reported intake and the measured expenditure using the doubly labeled water method is between 20% and 50% [17]. As a result, the lack of assessment tools capable of producing unbiased objective data have significantly hampered the progress of obesity research.

We believe that an effective solution to this problem can be found in the fields of electrical engineering and information technology. Rapid advances in these fields have yielded sophisticated devices and intelligent computational algorithms that automatically acquire and analyze multimedia data from the real-world environment. We propose the idea of a ubiquitous multimedia technology capable of providing an accurate estimation of intake energy based on food quantity. This technology can provide a powerful platform for the study of obesity.

A major difference between the proposed approach to dietary assessment and the existing approaches is the use of extensively and objectively recorded images or videos. With an electronic visual memory, this powerful technology stores all of the scenes that the individual has observed throughout the entire recording period. Since “seeing is believing”, our “monitoring” system allows for a more complete understanding of events than current non-video based methods.

1.2 EXISTING IMAGE-BASED APPROACHES TO FOOD DIMENSIONAL ESTIMATION

It has been shown that food size, while an important parameter, is difficult to accurately quantify by self-report or questionnaire. Several image-based approaches and software solutions have been proposed to monitor and manage dietary intake. In order to reduce the burden of response on the part of the user, electronic devices with built-in cameras have been proposed to simplify the self-monitoring process and increase computational accuracy, similar to a miniature camera or smart phone. In addition, it has been shown that technology-assisted logbook techniques are perceived as less intrusive upon lifestyle than traditional methods for recording dietary intake [18]. Therefore, approaches based on computer vision techniques for food volumes estimation have been developed in recent years. Approaches using a camera with a fixed position to estimate the volumes of fruits such as watermelon, kiwifruit and orange were reported by Koc [20], Rashidi [22] and Khojastehnazh et al [21]. However, these approaches require the relative positions of the camera and the object to be known in advance, and are only effective for certain food items. Spherical objects [23, 24] and checkerboard patterns [25] have been used as fiducial markers to reconstruct 3D objects from 2D images using computer vision techniques. Puri et al presented a system using three images and a checkerboard to estimate food volume [27]. The relative camera poses among the three images were estimated using a *peremptive* RANSAC-based method [26] and the scale ambiguity was determined by placing a checkerboard beside the food. Figure 2 shows the data flow of their system, through which they were able to achieve an accuracy rate of

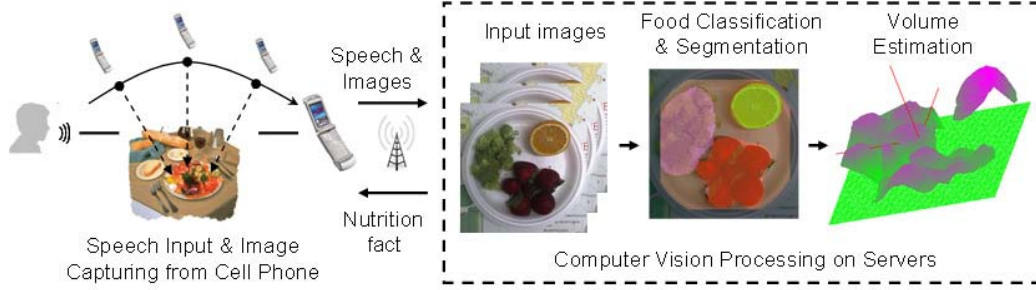


Figure 2: Data flow of the entire system in [27]

approximately 90%. A similar approach was proposed by researchers at Purdue University [19]. They used a checkerboard pattern in each food image to obtain camera parameters and provided a reference for the food scale (shown in Figure 3). Their testing objects consisted of spherical and prismatic objects only and their relative errors were between 7% to 15%. These methods required a relatively inconvenient procedure including placing a checkerboard at a position near the food before taking pictures. Moreover, the requirement to take several pictures before and after eating interferes greatly with normal eating patterns and the user’s lifestyle.

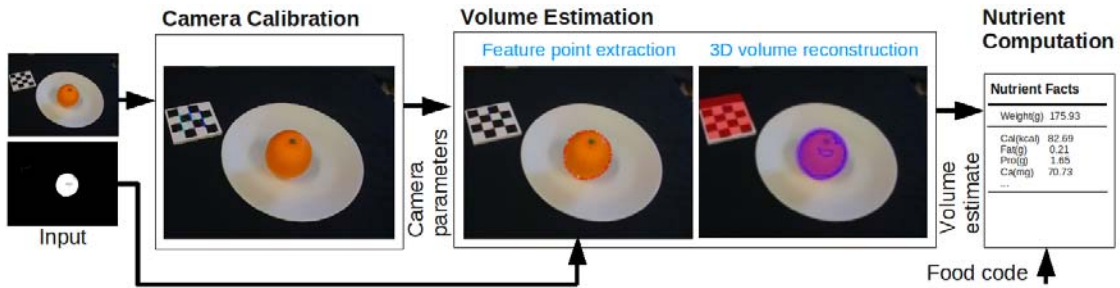


Figure 3: Process of the food portion estimation in [19]

This dissertation aims to overcome the problems of the existing systems. Our system will provide a set of image processing tools, which computationally measure the volume of food or drink automatically. Eliminating the need to carry a reference object, our proposed approach provides a minimum-effort user interface. Unlike other commercial and research

devices with image/video acquisition functions, the proposed device only requires a single miniature camera and simple structured lights, which are specially designed to precisely measure food quantity. Such designs allow greater cost, space and power savings.

This research designs geometric mathematical models and algorithms for our newly developed wearable devices to efficiently and accurately estimate food dimensions. Our techniques and algorithms can also be used in other wearable image acquisition devices to reconstruct the dimensions of indoor objects. The advantages of our techniques are:

- Specific designs of assistant structured lights with low power consumption;
- Low-cost and easy manufacturing, calibration, and maintenance;
- Estimation methods using a single image containing the objects of interest;
- Low complexity geometric algorithms for real-time application;
- Volume estimation for regularly shaped objects.

1.3 OTHER POTENTIAL APPLICATIONS OF PROPOSED DIMENSION ESTIMATION TECHNIQUES

As a mobile multimedia data acquisition device, our proposed system and techniques have many potential applications besides food volume measurement. One of the potential applications is the *electronic chronicle*, also called *e-chronicle*, which is defined as an extensive record of events obtained using multiple sensors and sources of information. An e-chronicle provides access to such data at multiple levels of granularity and abstractions along temporal and other contextual dimensions, and uses appropriate access mechanisms in representations and terminology familiar to application users [28]. For instance, smart phones, one of many forces driving the emergence of e-chronicles, are becoming an inseparable part of our lives as true multimedia devices that combine communications, computing, and contents [29]. The sensing and communication capabilities of mobile devices, coupled with e-chronicling technology, have the potential to change the everyday practices of people in both personal life and business activity [28]. Unlike other traditional e-chronicles systems which can only replay the recorded events manually, ours is an advanced system with the unique ability to

reconstruct 3D information of objects of interest. This will allow e-chronicles to be used in navigation, 3D retrieval, 3D visualization, etc.

Another potential application is virtual modeling, which is used extensively in computer games [30] and film making [31, 32]. After new objects are inserted and more view-points generated, as well as the basic indoor background scenes set up and precisely modeled, reconstruction of the backgrounds using video and stereo images is often required. However, 3D models using single-video-based methods usually lack precision or can only be reconstructed up to a scale. Though stereo systems can improve the scale ambiguity of reconstruction, they require more expensive equipment, testing, and calibration. The proposed techniques provide an alternative and easier way to reconstruct relatively simple background scenes, such as walls, tables, furniture, and floors for 3D computer games and 3D movie applications.

Our system and methods are also potentially useful in applications such as human face and body modeling, non-contact measurement, real-time object capturing, and 3D modeling and visualization.

1.4 CONTRIBUTION OF THIS DISSERTATION

This dissertation presents a practical development and implementation of a new food dimension estimation method. Our method provides a powerful platform for the study of obesity as well as other potential applications.

The original contributions of the dissertation are summarized as follows.

1. Specific structured lights designs based on laser and LED beams are developed and implemented for food dimension estimation. These designs are safe for human use, low in cost and power consumption, and are space saving. To the best of our knowledge, this is the first quantitative dietary assessment system using structured lights;
2. Geometric dimension estimation algorithms are developed and implemented. Both ideal and modified system models and algorithms are presented. Height estimation method is proposed for regularly shaped objects. Our algorithms have low computational complex-

ity and only need a single image for estimation, which is especially suitable for mobile and real-time applications.

3. A practical structured lights calibration approach is developed and implemented. An orthogonal linear regression method is presented to fit spatial lines. The proposed approach can quickly and easily calibrate our camera and structured light system and greatly improve the estimation accuracy.
4. Color and shape characteristics of reference features in the captured images are investigated. Segmentation methods using a histogram-based thresholding are proposed for both laser and LED spots to extract structured light patterns automatically. An ellipse fitting method is implemented to improve the accuracy in feature extraction.
5. The proposed approach is tested in experiments using various artificial items and real foods. The average relative food size estimation error is less than 10% for the laser beams design, which is much better than the self-reporting and questionnaire methods. The effects of image resolution, position and orientation of objects and the shape of objects on the proposed algorithms are further investigated. Our approach gives better estimation accuracy when an object's orientation is between 20° and 50° .

2.0 BACKGROUND

In this chapter, we review the basic concepts of pinhole camera geometry including perspective projection and affine projection, and introduce definitions of the intrinsic and extrinsic parameters of the camera. We also review 3D reconstruction techniques. The problem of dimensional reconstruction from one single image is described, and an extensive review of existing methods is presented.

2.1 BASIC CONCEPTS OF THE PINHOLE CAMERA MODEL

There are many types of imaging devices that convert measurements of light into information about spatial and material properties of a scene, such as a camera and a retina. The first model of the camera invented in the 16th century used a pinhole to focus light rays onto a wall or translucent plate, which demonstrated the laws of linear perspective (shown in Figure 4 (a)) discovered a century earlier by Filippo Brunelleschi [47]. Later, the pinholes were replaced by more complicated lenses.

In current computer vision community, people usually use the mathematical model of a pinhole lens or a thin lens instead of a real optical system. The principle of a pinhole camera is shown in Figure 4 (b). Light rays from an object pass through a small hole to form an image. These models not only reduce the process of image formation by tracing rays from points on objects to pixels in images, but also are proved to be workable in most applications [46]. There are two major geometrical projections used for pinhole camera models: perspective projection and affine projection. We will introduce them in the following section.

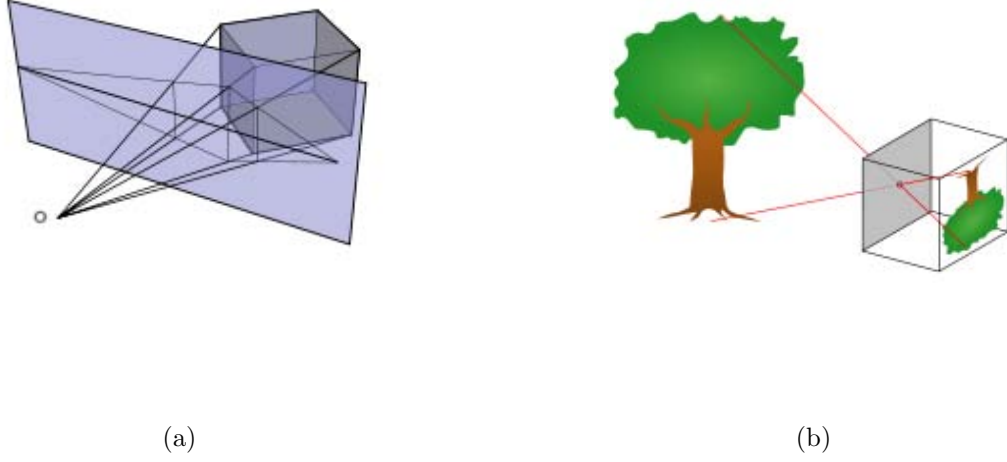


Figure 4: (a) Rays of light travel from the object through the picture plane to the viewer's eye O . (b) Space principle of a pinhole camera. (Courtesy of [47])

2.1.1 Perspective and Affine Spaces

A frontal pinhole imaging model is shown in Figure 5. The image of a 3D point P is the point p , which is the intersection of the image plane and the ray passing through the pinhole camera optical center O . f is the focal length in front of the optical center.

The *optical axis* of a camera is an imaginary line that defines the path along which light propagates through the system. For a system composed of simple lenses, the optical axis passes through the center of curvature of each surface, and coincides with the axis of rotational symmetry [47]. In a pinhole camera model, the *camera plane* is defined as the imaging sensors plane that contains the object's projected image, and lies beyond the focal point.

Let us consider a generic point P in 3D space with coordinates $P = [x_c, y_c, z_c]^T \in \mathbb{R}^3$ in the camera coordinate system (also called camera frame). The same point has coordinates $P = [X_w, Y_w, Z_w]^T$ in the world coordinate system (also called world frame). The rigid-body transformation between the two coordinates' systems can be described by a rotation matrix

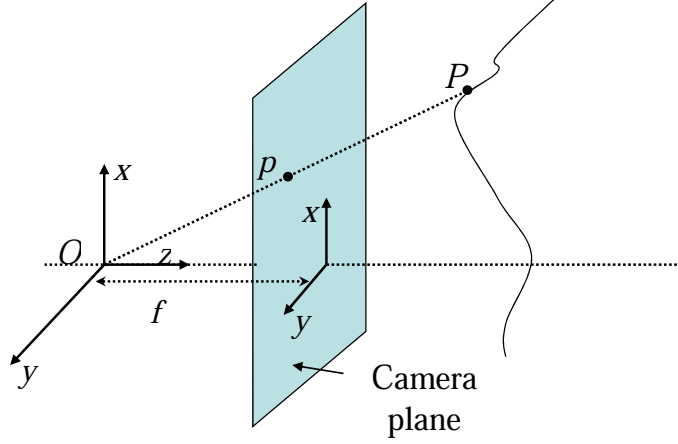


Figure 5: Frontal pinhole imaging model

$\mathbf{R} \in \mathbb{R}^{3 \times 3}$ and a translation vector $\mathbf{T} \in \mathbb{R}^{3 \times 1}$. Therefore we can establish the relationship between the coordinates of a point under two coordinates frames as follows:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{T} \quad (2.1)$$

The rotation matrix \mathbf{R} is characterized by the following properties: (1) the inverse of a rotation matrix is equal to its transpose, i.e. $\mathbf{R}^T \mathbf{R} = \mathbf{R} \mathbf{R}^T = \mathbf{I}$; and (2) its determinant is equal to 1, i.e. $\det(\mathbf{R})=1$. By this definition, the columns of a rotation matrix form a right-handed orthonormal coordinate system, so do its rows.

Adopting the pinhole camera model in Figure 5, we get the *perspective projection* equation, which projects a point $P(x_c, y_c, z_c)$ onto the camera plane at $p(x, y)$

$$p = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{z_c} \begin{bmatrix} x_c \\ y_c \end{bmatrix} \quad (2.2)$$

where f is the focal length and z_c is the P 's coordinate in z direction.

As noted, pinhole perspective is only an approximation of the geometry of the imaging process. Other approximations such as *affine projection* models are used in some applications.

One of the affine models, the *weak-perspective* model, assumes that all line segments in the fronto-parallel plane are projected with the same magnification. Therefore the projection equation can be rewritten as

$$\begin{aligned}x' &= -mx \\ y' &= -my\end{aligned}\tag{2.3}$$

where $m = -\frac{f}{z}$ and z is scene depth to the pinhole.

With the assumption that the camera always remains at a roughly constant distance from the scene with all light rays parallel to the camera optical axis and orthogonal to the image plane, we can further normalize the image coordinates and let $m = -1$. This is called *orthographic projection*, which is defined by

$$\begin{aligned}x' &= x \\ y' &= y\end{aligned}\tag{2.4}$$

Although weak-perspective projection is an acceptable model for many imaging conditions, the assumption of pure orthographic projection [33] is usually unrealistic. In this work, we employ the most frequently used perspective projection model in all the algorithms and performance analysis.

2.1.2 Camera Intrinsic and Extrinsic Parameters

In computer vision, the intrinsic and extrinsic parameters of the camera are used to describe the mapping from the coordinates of a 3D point to the 2D image coordinates of the point's projection onto the image plane. Therefore, we introduce these two important definitions.

2.1.2.1 Intrinsic Parameters

To specify the relationship between the camera plane coordinate system and the pixel array in an image, we consider the model shown in Figure 6. Let (x, y) be metric units (e.g. millimeters), and (x_s, y_s) be scaled versions that correspond to the coordinates of the pixel.

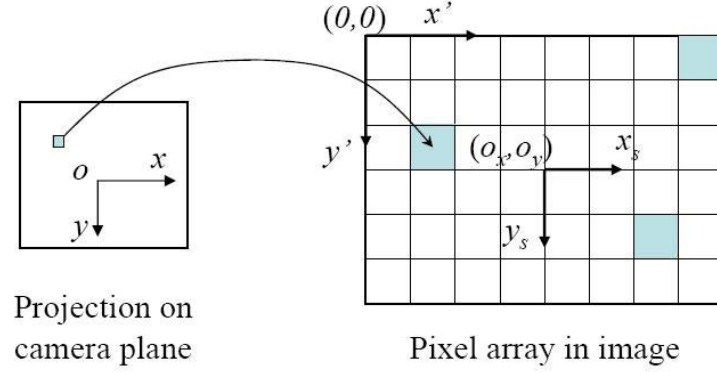


Figure 6: Image coordinate frame origin translation.

The transformation can be described by a scaling matrix

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.5)$$

Here s_x and s_y depend on the size of the pixel in metric units (sometimes called *pitch size*) along the x and y directions.

In practice, a digital camera records the measurements in terms of pixels (x', y') with the origin of the image frame typically in the upper-left corner of the image. We need to translate the origin of the reference frame to this corner as shown in Figure 6.

$$\begin{aligned} x' &= x_s + o_x \\ y' &= y_s + o_y \end{aligned} \quad (2.6)$$

where (o_x, o_y) are the coordinates (in pixels) of the principal point relative to the image reference frame.

In cases where the pixels are not rectangular, a more general form of the scaling matrix can be considered

$$\begin{bmatrix} s_x & s_\theta \\ 0 & s_y \end{bmatrix} \in \mathbb{R}^2 \quad (2.7)$$

where s_θ is called a *skew factor* and is proportional to $\cos(\theta)$, where θ is the angle between the image axes x_s and y_s .

The coordinate transformation matrix (also called *intrinsic matrix*) in homogeneous representation has the following general form

$$\mathbf{K} = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^3 \quad (2.8)$$

Thus, the homogenous coordinate transformation between a point on the camera plane and its image is as follows:

$$\begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.9)$$

2.1.2.2 Extrinsic Parameters

The geometric relationship between a point of coordinates $X = [X_w, Y_w, Z_w]$ relative to the world frame and its corresponding image coordinates $x = [x', y', 1]$ (in pixels) depends on both the intrinsic matrix \mathbf{K} and the rigid-body motion (\mathbf{R}, \mathbf{T}) relative to the camera frame, which are called *extrinsic calibration parameters*. The camera extrinsic parameters describe the coordinate system transformations from 3D world coordinates to 3D camera coordinates. Therefore, with the intrinsic matrix \mathbf{K} and extrinsic parameters \mathbf{R} and \mathbf{T} , we can model the image formation with a general *projection matrix* $\mathbf{P} = [\mathbf{KR}, \mathbf{KT}]$. It is of practical interest to put some restrictions on the intrinsic parameters since some of these parameters are fixed and known. A camera with a known nonzero skew and non-unit aspect-ratio can be transformed into a camera with zero skew and unit aspect-ratio by an appropriate change of image coordinate. Faugeras proved the necessary and sufficient conditions for a perspective projection matrix [48]. Here we ignore the linear and nonlinear radial distortions by assuming that we can eliminate those with pre-processing. The interested reader can refer to related literature such as [49, 50, 55] for details.

To find the camera’s intrinsic and extrinsic parameters, a process called *Camera Calibration* is often implemented at an early stage in computer vision. Since the accuracy of those parameters is critical, research has been conducted to find robust and efficient methods for camera calibration under different situations. We will introduce details of our camera calibration and structured light calibration approaches in Chapter 5.

2.2 REVIEW OF 3D RECONSTRUCTION METHODS

Reconstruction of real world objects has many applications, from archaeology to architecture, and even more emerging areas like the biomedical and film industries. There are several popular techniques and research areas which use various equipment and methods to achieve similar goals. We briefly introduce some common methods in this section.

2.2.1 3D Reconstruction from Range Scanner

Range scanners can densely measure the distance from sensors to reflective points on objects in a grid pattern using points, strips, or 2D patterns. If we represent the range of each point as pixel intensity, the recorded grid pattern can be visualized as a range image. To construct a complete 3D model, range data from multiple views is required. There are a variety of technologies for acquiring the range of points. They are typically divided into two types: contact and non-contact 3D scanners.

Contact 3D scanners probe the subject through physical touch. A CMM (Coordinate Measuring Machine) is an example of a contact 3D scanner. The disadvantage of CMMs is the risk of modifying or damaging the measuring object. They are also relatively slow compared to other scanning methods. Non-contact 3D scanners can be further divided into two main categories, *active scanners* and *passive scanners*. An active scanner emits some type of radiation or light and detects its reflection in order to compute the structure. Visible or invisible light, such as ultrasound and x-ray are some common forms of emission. Based on their physical models, various methods have been proposed to use time-of-flight,



(a) stereoscopic camera



(b) stereoscopic viewer

Figure 7: Stereoscopic camera and stereoscopic viewer

triangulation, structured or modulated light, and Computed Tomography technologies for active scanners. Non-contact passive scanners do not emit radiation themselves, but instead work by detecting the reflected ambient radiation. Both visible light and infrared can be used for such scanners. One advantage of passive methods is that they are inexpensive, such as a common digital camera [73]. [60] describes a few of the common range scanning technologies and walks through a pipeline taking the range data into a single geometric model. [61] proposes a 3D mapping of indoor and outdoor environments using a mobile range scanner method, which smooths surface by area decreasing flow. Recently, Song Zhang and Peisen Huang from Stony Brook University developed a real-time scanner using digital fringe projection and a phase-shifting technique (a various structured light method). The system is able to capture, reconstruct, and render the high-density details of dynamically deformable objects (such as facial expressions) at 40 frames per second [62]. [73] reviews the developments in the field of 3D laser imaging in the past 20 years, and emphasizes commercial techniques and systems currently available.

2.2.2 3D Reconstruction from Stereo or Successive Image Sequences

The motivation for stereo methods comes directly from the biological structure of human eyes. These technologies are non-contact and passive, and the depth from binocular disparity

provides the fundamentals of all variations of stereo reconstruction methods. Stereoscopic image pairs were a popular form for education and entertainment in the late 19th and early 20th centuries[47]. These pictures were taken with a camera (much like the one in Fig 7 (a)), which had two lenses mounted a few inches apart. A stereoscopic viewer (Figure 7 (b)) was used to view the pictures which were mounted in pairs on cards a few inches apart so that they created a three dimensional effect. There is a huge amount of work investigating each major component of this problem, such as stereo camera calibration [63,64,67], epipolar geometry [65], essential matrix and fundamental matrix acquisition [46,48], correspondence matching [46,48,66], image rectification[68–71], disparity computation [72], and so on. Active stereo with structured light is also used to simplify the correspondence problem [76].

The stereo approaches require accurate calibration of the camera optical parameters and physical location. However, in many applications the information is not available or reliable enough. Autocalibration or *self-calibration* from motion provides a solution to reconstruct metric structure from video sequences in the uncalibrated case. Hartley and Zisserman provided a useful discussion of the self-calibration methods, implementations and evaluations in [46] for solving this kind of problem.

2.2.3 Structured Lights Aided 3D Reconstruction

The principle of using structured light is that projecting a certain pattern of light on to a shaped surface produces a distorted illumination which can be used for an exact geometric reconstruction of the surface shape. Structured light methods utilize active range sensing, which helps to develop highly accurate correspondence algorithms to reconstruct a precise 3D structure.

One of the most common forms of structured light is *optical triangulation*. Figure 8(a) shows its fundamental principle [60]. A focused beam of light illuminates a tiny spot on the surface of an object. This light is scattered in every directions for a matt surface. We can compute the position of the center point of this spot and trace a line of sight through that point until it intersects the illumination beam at the point on the surface of the object. If we use a plane of light instead of a beam, as shown in Figure 8(b), we can sweep the light

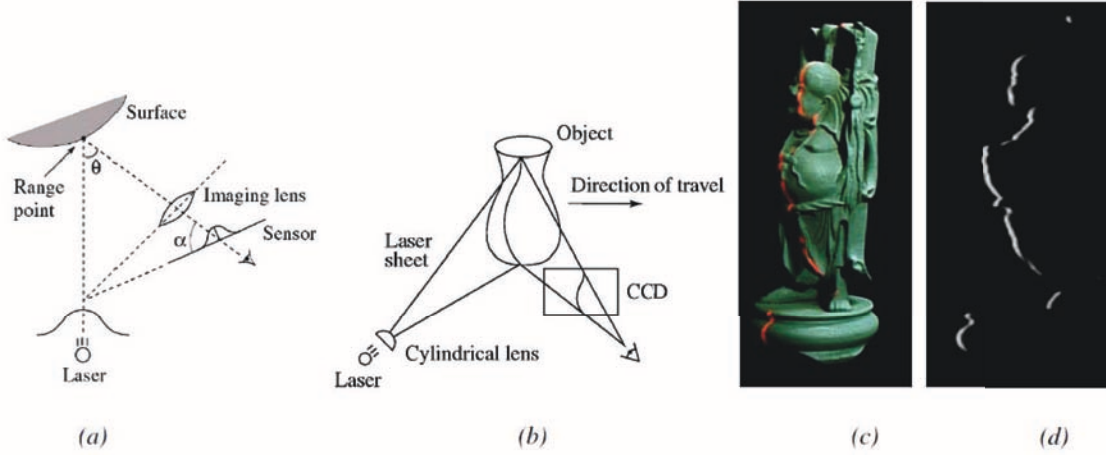


Figure 8: Optical triangulation. (a) 2D triangulation using a laser beam for illumination; (b) Extension to 3D; (c) Red laser line projected onto small (20cm) statuette; (d) Reflected light captured by CCD camera. (Courtesy of [60])

over the surface of the object and compute its shape based on the intersection of the line of sight with the laser plane. Figure 8(c) and (d) show a light stripe cast onto a real object and the reflection captured by the camera. Other techniques such as multi-strip system and hierarchical striping methods are proposed to reduce the number of shots [74]. Although many variants of structured light projection are proposed, patterns of parallel stripes are widely used. Figure 9 shows a parallel stripes projection on a seat and its reconstruction [47]. Imaging radar, based on time of flight radar system, is becoming increasingly popular because of its accuracy and ease of use [75]. [76] proposed a complex system that can be used to set up a ground truth data base for testing performance of different 3D reconstruction methods.

2.2.4 3D Reconstruction from a Single Image

It is well known that, a single image provides less geometric cues to infer depth than stereopsis and multiple images. There are a few active research efforts contributing to reconstruction of 3D information from a single image such as shape from shading and texture, vanishing



Figure 9: 3D reconstruction of a car seat (Photo courtesy of [47])

points algorithms, statistical model based methods, focusing or defocusing methods, structured lights methods, circular feature methods, and approaches using other assistant optical instruments. Since we use a single camera to reconstruct 3D structure in our proposed device, we will briefly review these aforementioned methods in more details in Section 2.3.

2.3 ALTERNATIVE METHODS FOR 3D RECONSTRUCTION FROM A SINGLE IMAGE

2.3.1 Shade from Shading and Texture

Shape from shading (SFS) methods basically recover the shapes from a gradual variation of shading in the image. The Lambertian model [58] in Figure 10 has been used widely as a simple model of image formation, in which the pixel gray level depends on the light source direction and the surface normal. However, real images do not always follow the Lambertian model. Many pioneers researched reconstruction for decades to use shading of one view [34, 35, 38]. Zhang et al. implemented and compared six well-known SFS algorithms in their survey [38], and concluded that, in general the minimization approaches are more robust, while other approaches are faster. Recently, a method was proposed in [37] for reconstructing the shape of a deformed surface from a single image by combining the cues of shading and

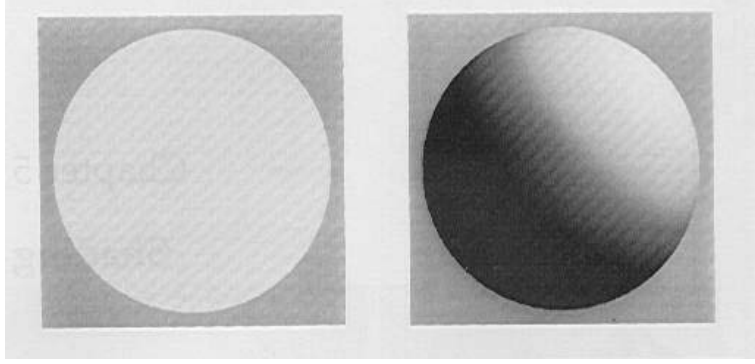


Figure 10: From a monocular view with a constant intensity light source (left figure) and a single distant light source of known incident orientation upon an object with known reflectance map (right figure)

texture. Their method solved the ambiguities locally but required a texture estimate as a priori knowledge. The quality of their reconstruction was a middle ground between the flexibility of single cue, single view reconstruction and the accuracy of multi-view techniques.

2.3.2 Vanishing Point

In [39, 40, 59], vanishing points (VP) were proposed to be used to reconstruct 3D shape from a single image. Those methods assumed perfect perspective projection in the images and used the lines drawn on an object as clues to rebuild their 3D structures. Yoon et al. present a method using VP to recover the dimensions of an object and its pose from a single image with a camera of unknown focal length [41]. However, their algorithm needed to accurately extract the information of a VP in order to have orthogonal property from the image, otherwise nonlinear optimization techniques were not able to improve the accuracy in comparison with conventional methods. Jelinek modeled the object as a polyhedron where a linear function of a dimension vector is expressed using the coordinates of the vertices [42]. This was also a nonlinear optimization problem and greatly relies on the initialization selection and searching strategy.

2.3.3 Statistical Models

Recently, Andrew Y. Ng and his group estimated detailed 3D structure from unstructured indoor and outdoor environments for both quantitatively accurate as well as visually pleasing purposes [77–79]. They combined the monocular image cues with triangulation cues to build a photo-realistic model of a scene. After learning the relation between the image features and the location/orientation of the planes, and also the relationships between various parts of the image, a hierarchical, multiscale Markov Random Field (MRF) was used to predict the value of the depth map as a function of the image. They believed that even a single image can infer a significant portion of the scene’s 3D structure [43]. Their algorithm investigated the visual cues for scene understanding and worked well for the visualization of large-scale scenes. However their reconstruction was only quantitatively accurate, which was insufficient for precise dimension reconstruction.

Another statistical image-based model proposed by Grauman [80] could infer 3D structure parameters using a probabilistic “shape+structure”. The 3D shape of an object class was represented by a set of contours from silhouette views simultaneously observed from multiple calibrated cameras. Given a novel set of contours, the method inferred the unknown structure parameters from the new shape’s Bayesian reconstruction. Their shape model enabled an accurate estimate of structure despite segmentation error or missing views in the input silhouettes, working with only a single input view. However, the proposed model was only based on human body silhouette structures which are not easily transferrable to other 3D objects.

2.3.4 Focusing and Defocusing Methods

The focusing and defocusing method is another major technique which can use only one image to reconstruct 3D dimensions. *Image Focus Analysis* (IFA) methods search the camera parameters that correspond to focusing the object [84, 85]. A larger number of images are needed as input to compute a focus measure in order to determine the focused image and 3D shape. Subbarao and Tae-Choi [83] proposed a shape from image focus methods which was based on finding the best focus measure on a focused image surface instead of over image

frames sensed by a planar image detector. The problem of how to choose the best focus measure was also an important issue for any IFA method. Some examples of focus measures are image energy, energy of Laplacian and energy of image gradient [86].

Image Defocus Analysis (IDA) has recently attracted the attention of researchers. This method is usually based on the assumption that a defocused image of an object is the convolution of a sharp image of the same object with a two-dimensional window function, for example Gaussian function, whose parameters are related to the object depth. Pentland and Subbarao [44, 93] firstly proposed to use the amount of defocus or blurring to determine its depth and Grossamnn called it *the depth-from-focus* (DFF) method [94]. Other algorithms based on the same basic assumption appeared later [81, 82, 87, 88]. The fundamental advantage of the newly developed methods is the two-dimensionality of the aperture, which is less sensitive to the noise disturbance of measurements and allows more robust estimation [89].

2.3.5 Structured Light Methods

Since structured light and devices provide greater depth cues, various instruments have been developed based on their requirements. A part from the technologies we introduced in Section 2.2.3, several specially designed lights or devices have been proposed to compensate for the shortcomings of a single image. Lu [90] split a single camera view into stereo vision, with an option to project a color-coded light structure onto the object using a synchronized flash light source. Levin [91] carried out their modification to insert a patterned occluder within the aperture of the camera lens, creating a coded aperture. Using a statistical model of images, they recovered both depth information and an all-focus image from single photographs taken with the modified camera. A fish-eye lens and a cylinder whose inside is coated by a silver reflective layer were developed and used in [36] to capture an image that includes a set of points observed from multiple viewpoints.

2.3.6 Circular Feature Methods

3D location estimation, which includes 3D position estimation and 3D orientation estimation, have been extensively addressed in the literature. They play important roles in 3D

scene reconstruction and robot navigation. With known camera intrinsic parameters and 3D location of the object, it is straightforward to reconstruct the 3D structure of an object. Circular features are the most common quadratic-curved features that have been addressed for 3D location estimation. The advantages of using a circular shape are that [97]: (1) many manufactured objects have circular holes or circular surface contours; (2) a circle has good properties from a mathematical perspective, such as its perspective projection in any arbitrary orientation is always an exact ellipse, and it can be defined with only three parameters due to its symmetry with respect to its center; (3) a circle has been shown to have the property of high image-location accuracy [95, 96]; and (4) the complete boundary or an arc of a projected circular feature can be used without knowing the exact point correspondence.

For circular feature-based 3D location estimation, approximation-based methods [98–100] and closed-form solution methods [101–103] have been proposed. In the approximation-based methods, difficulties remain for the mathematical presentation. In order to reduce the complexity of a 3D problem, some simplifying assumptions have to be made, such as assuming that the optical axis of the camera passes through the center of the circular feature, using affine or orthogonal projection instead of perspective projection, etc. Moreover, it is very difficult to analyze the error introduced by each approximation in some methods. There have been several methods proposed in a scene under general conditions without simplifying assumptions [101, 102]. However, these methods are mathematically complex and algebra-based. They lack geometric representations of the problem and geometric interpretations of the solutions.

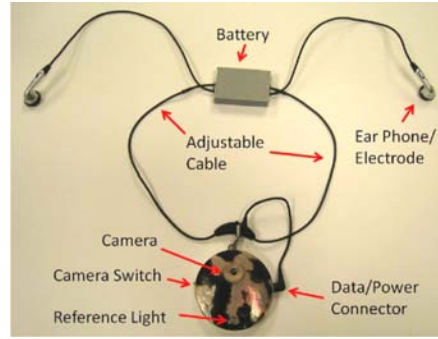
Various optical assistant systems have been presented to measure three dimensional structure. Unfortunately, most of the previous structured light methods need either a well trained model or complicated assistant system, which greatly increase the cost of implementation. In this dissertation, we present new device designs with simple assistant structured light and fast algorithms. Although our designs relating to food dimension estimation were motivated by a specific target application, they can be used in other applications, such as mobile consumer electronics, health care devices, and robotic navigation.

3.0 MEASUREMENT SYSTEM DESIGN

3.1 DESIGN OF WEARABLE DEVICE FOR DIETARY ASSESSMENT



(a) Subject wearing the electronic mobile device



(b) Close view of the wearable device

Figure 11: Prototype of wearable electronic device

Recent advances in microelectronics offer unique opportunities to develop electronic devices and methods for objective dietary assessment. Our group is leveraging these advances to design and evaluate new systems that minimize intrusion into the users' lives. Currently, we have designed and implemented several prototypes and are still working to improve the accuracy and mobility of these devices and their supporting softwares.

The electronic system contains two major components, a wearable unit and a data analysis software package installed in a computer at the dietitians office. One of the prototypes of the wearable unit is shown in Figure 11. The circular device presently looks like a wear-



Figure 12: Laser point-array pattern generated by a diffraction grating glass

able MP3 player, suitable for both men and women. The appearance of the device may be tailored according to individual preference. Currently, the device has a diameter of 62 mm, which will be reduced substantially in the future. A large-capacity (2700 mAh), rechargeable, lithium-ion battery (rectangular object in Figure 11) is placed in the back of the neck connected to the circular device using adjustable cables.

As a significant component of our mobile device, the development and implementation of a 3-dimensional estimation unit using a miniature camera with reference lights is the major focus of in this dissertation. A safe-to-human structured light pattern is needed in our dimensional estimation unit to provide 3D information for retrieving the dimensions of food objects. Low-power laser modules and LED diodes are two appropriate candidates of the structured lights which are described in the following sections.

3.2 LASER BEAM BASED STRUCTURED LIGHT DESIGN

In one of our prototypes, we designed a dimension estimation unit using laser beam based structured lights with a single camera. With the help of a pre-measured laser beam system,

we were able to estimate the physical dimensions of objects on an arbitrarily positioned plane. We investigated a method using one laser module with an attached diffraction grating film to project a point array.

The major advantage of using one laser module with a diffraction film is the reduced costs of both space and energy. The diffraction grating film diffracts light into several beams traveling in different directions. The directions of these beams depend on the spacing of the grating and the wavelength of the light. As a spatially coherent (with identical frequency and phase), narrow beam, the laser light is diffracted by the grating film and forms a point-array with a known pattern. Figure 12 shows a diffraction grating glass creating an expanding grid of multiple laser beams when a single laser beam is passed through it. However, the energy of the original beam will be split after the diffraction. The farther the points are located from the center of the original beam, the weaker the intensity at these points. Under common illumination conditions, the outer points are difficult to be detected by a camera, which will bring difficulty for feature detection in the image processing phase.

Using more laser modules can help to solve the visibility issue of the features. To determine the location of a 3D plane, we need at least three noncollinear points. Therefore, we proposed an original method to use three individual laser modules to construct a structured triangle as a reference pattern with the allowance of system energy distribution. Figure 14(a) shows a testing board with three laser highlighted spots.

For safety considerations, we chose low intensity optical diodes to protect human eyes. An experimental model shown in the left image in Figure 13 was used to verify and evaluate the proposed dimension estimation algorithms. Three laser diodes were mounted on a plexiglass board. Each laser diode and a focusing lens were adhered to a small panel which was attached to curved sheets of copper as shown in the right image in Figure 13. A screw was used to adjust the beam angle by forcing the copper sheet to bend. A voltage/current source was used to drive the laser diodes. The three diodes were connected in series. The operational voltage and current of each of the three laser diodes were 2.1V and 20mA, respectively, and the operational power of each laser diode was about 40mW.



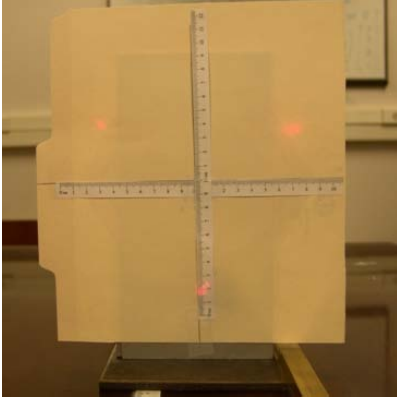
Figure 13: Experimental instruments with three laser diodes

3.3 LED BASED STRUCTURED LIGHT DESIGN

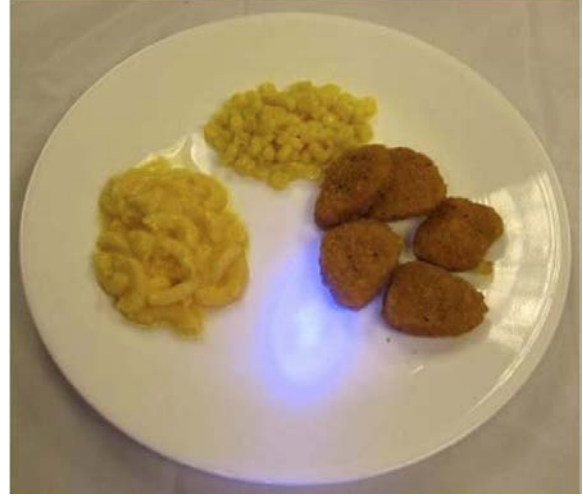
In order to minimize the use of power and space, we developed another structured light system, which uses only one LED diode. We also studied its corresponding geometric algorithm to estimate the dimensions of the spatial object. The LED diode produced an ellipse-like pattern on the object (Figure 14(b)) which provided enough information to locate the position and orientation of the object plane. Our algorithm was based on the circular feature perspective projection geometry between the real world and a camera plane to estimate an object's dimension using its position information.

3.4 SYSTEM INTEGRATION

It is critical for our prototype system to be portable and functional enough for real-life application. As shown in Figure 11, two separate smaller devices, connected to the front device by the same cables, reach locations on the upper body and head. These devices can be earphones, microphones, accelerometers, or skin-surface electrodes for various applications, such as user interface and physiological measurements. The front device contains several sensors and data processing/storage components, including a miniature camera for video recording,



(a) Object with three laser highlighted spots



(b) Food with LED spot

Figure 14: Structured lights with object of interest

reference lights for food dimension measurement as described above, an accelerometer for physical activity monitoring, a microphone for the identification of eating episodes using ambient sound, a global positioning system (GPS) for location identification, a central processor for data processing, and an 8-GB microSDHC flash memory for storing data. All electronic components are installed on an eight-layer printed circuit board. The hardware design has been completed, and the embedded software (routines that provide interface to hardware modules) is being developed.

4.0 GEOMETRIC DIMENSIONAL ESTIMATION ALGORITHMS BASED ON PERSPECTIVE AND STRUCTURED LIGHTS

To measure an object's physical dimensions from a single still image is a difficult when only one view of a scene is available. Most of the solutions for single image 3D reconstruction problems must be based on assumptions of the object's shape and the knowledge of explicit camera intrinsic parameters, camera position, and the location of the object of interest. However, the information is difficult to obtain in most cases, especially for a mobile device. In this chapter, we propose several new measuring methods utilizing our structured lights designs, using either laser beam or LED spotlight. Using the geometric information of the pre-measured structured lights, we can estimate the physical dimensions of objects on an arbitrarily positioned plane.

We mainly focus on two canonical types of measurement: (i) dimensions of segments on planar surfaces and (ii) distances of observation points to the object of interest. In many cases, these two types of measurements have been proved to be sufficient for a partial or complete three-dimensional measurement and reconstruction of the observed scene.

4.1 ASSUMPTIONS AND GEOMETRIC RELATIONS AMONG IMAGE, CAMERA PLANE, LASER BEAMS AND OBJECT PLANE

Figure 15 shows the geometric model of our measurement system. In this figure, I is one of the three laser beams, and $O'A'$ is the perspective projection ray of one laser highlighted spot A' on an object plane. O is the pre-determined origin of the world coordinate system. The plane which is defined by the triangle ABC is the camera plane. The center of the

camera plane is on the optical axis passing through the camera optical center O' . 2β is the camera angle of view and z_0 is the focal length. A', B', C' are three laser highlighted spots on the object plane and A, B, C are their projections on the camera plane. The transformation between camera plane coordinate system and world coordinate system has been described in Eq. 2.1.

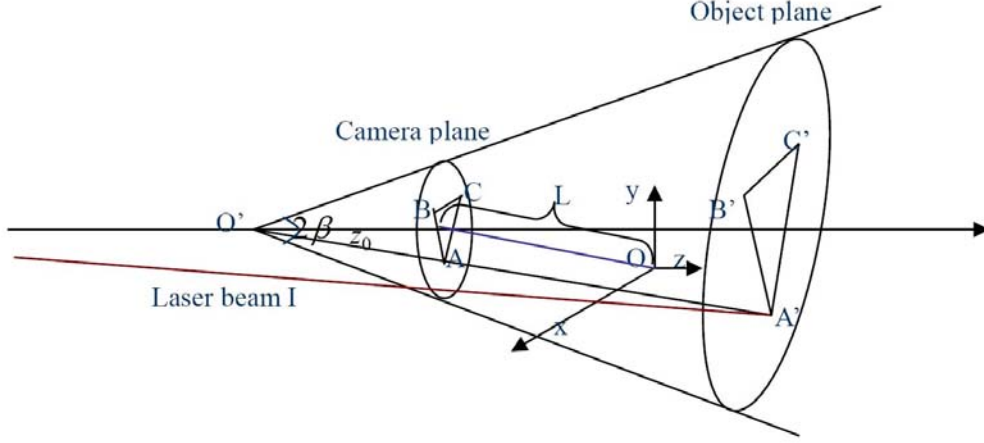


Figure 15: Geometric relationships among camera plane, object plane, and camera origin

The point A' on the object surface is the highlighted spot caused by the reflection of laser beam I . Two skew lines I and $O'A'$ intersect at point A' . Under ideal conditions, if the equations of two spatial lines are exactly known, we can solve them and obtain the coordinates of A' . Therefore, in order to obtain the coordinates of points A', B', C' , we need to know equations of six lines and to calculate the intersection of each pair of lines, including line I and line $O'A'$, line II and line $O'B'$, and line III and line $O'C'$.

4.2 IDEAL LASER BEAM MODEL AND CORRESPONDING ALGORITHM

We utilize a pinhole camera perspective projection model (described in section 2.1.1) to reconstruct the optical path of points of interest. The three laser beams are calibrated using

an image processing approach and a fitting method called *Orthogonal Linear Regression* which will be described in Chapter 5. Based on perspective geometry, we estimate the equation of an arbitrarily positioned plane, where the points of interest are located. Once the equation of this plane is successfully estimated, we can estimate the dimensions of any objects on the determined plane.

An ideal setup of three laser beams is shown in Figure 16. The three laser modules are symmetrically mounted at position $O'(0, 0, -L)$ and their beams can construct an equilateral triangle on a plane perpendicular to the camera optical axis (z axis). Other parameters are the same as those described in Section 4.1. The angles between each ray of light and the z axis is α . If we fix the angle between beam I and the x axis at $\alpha + \pi/2$, the angle between I and y axis is $\pi/2$. It is not difficult to calculate the angles between the other laser beams with x and y axes as:

$$\begin{aligned} \text{Angle between II and x axis: } & \arccos\left(\frac{\sin \alpha}{2}\right) \\ \text{Angle between II and y axis: } & \arccos\left(\frac{\sqrt{3} \sin \alpha}{2}\right) \\ \text{Angle between I and x axis: } & \arccos\left(\frac{\sin \alpha}{2}\right) \\ \text{Angle between I and y axis: } & \pi - \arccos\left(\frac{\sqrt{3} \sin \alpha}{2}\right) \end{aligned}$$

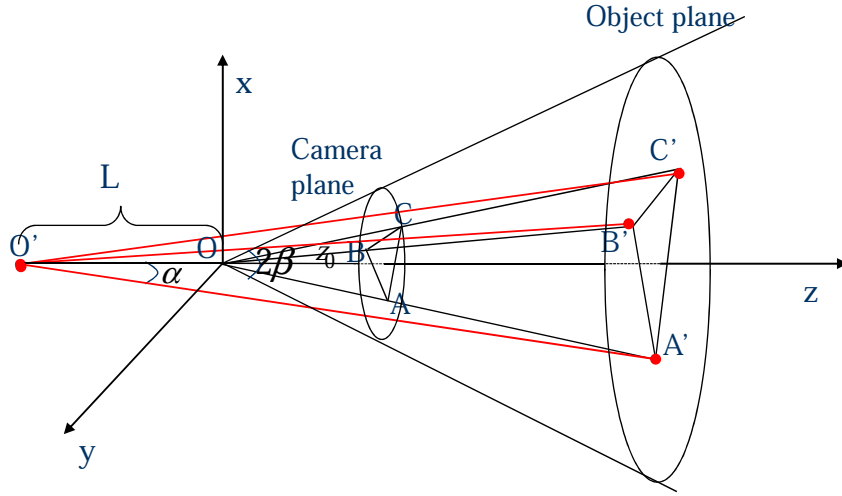


Figure 16: Geometric relationship of ideal laser beams design

Without loss of generality, we assume that the coordinates of A, B, C on the camera plane are $A(x_A, y_A, z_0)$, $B(x_B, y_B, z_0)$, $C(x_C, y_C, z_0)$. According to the two-point form of a

line equation, we can obtain the spacial equations for three virtual perspective projection rays as

$$\begin{aligned} OA : y = 0, \frac{x}{x_A} &= \frac{z}{z_0} \\ OB : \frac{x}{x_B} &= \frac{y}{y_B} = \frac{z}{z_0} \\ OC : \frac{x}{x_C} &= \frac{y}{y_C} = \frac{z}{z_0} \end{aligned} \quad (4.1)$$

By applying the point-slope form, with point $O'(0, 0, -L)$ and skew line direction vectors $[\cos \pi/2, \cos(\alpha + \pi/2), \cos \alpha]$, $[\sqrt{3} \sin \alpha/2, \sin \alpha/2, \cos \alpha]$, $[-\sqrt{3} \sin \alpha/2, \sin \alpha/2, \cos \alpha]$, we can write the line equations of $O'A', O'B', O'C'$ as

$$\begin{aligned} O'A' : y = 0, \frac{x}{-\sin \alpha} &= \frac{z+L}{\cos \alpha} \\ O'B' : \frac{x}{(\sin \alpha)/2} &= \frac{y}{(\sqrt{3} \sin \alpha)/2} = \frac{z+L}{\cos \alpha} \\ O'C' : \frac{x}{(\sin \alpha)} &= \frac{y}{-(\sqrt{3} \sin \alpha)/2} = \frac{z+L}{\cos \alpha} \end{aligned} \quad (4.2)$$

Note that point A' is the intersection of the corresponding spatial lines OA and $O'A'$, so are B' and C' . Therefore, coordinates of the three points, A' , B' , and C' , can be obtained by solving the equations

$$\begin{aligned} OA &= O'A' \\ OB &= O'B' \\ OC &= O'C' \end{aligned} \quad (4.3)$$

Thus the solutions are

$$\begin{aligned} A' : & \left[(x_A, 0, z_0) \times \frac{-\tan \alpha}{x_A + z_0 \tan \alpha} L \right] \\ B' : & \left[(x_B, y_B, z_0) \times \frac{\tan \alpha}{2x_B - z_0 \tan \alpha} L \right] \\ C' : & \left[(x_C, y_C, z_0) \times \frac{\tan \alpha}{2x_C - z_0 \tan \alpha} L \right] \end{aligned}$$

We can then compute the equation of plane $A'B'C'$ by setting a parametric equation given by

$$c_1x + c_2y + c_3z = 1 \quad (4.4)$$

After substituting the coordinates of A', B', C' in Eq. 4.4, respectively, and rearranging them to a matrix form, we have

$$\mathbf{T}\mathbf{c} = \mathbf{b}, \quad (4.5)$$

$$\text{where } \mathbf{T} = \begin{bmatrix} -x_A & 0 & z_0 \\ x_B & y_B & z_0 \\ x_C & y_C & z_0 \end{bmatrix}, \text{ undetermined coefficients } \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} \frac{x_A + z_0 \tan \alpha}{L \cdot \tan \alpha} \\ \frac{2y_B - z_0 \tan \alpha}{L \cdot \tan \alpha} \\ \frac{2y_C - z_0 \tan \alpha}{L \cdot \tan \alpha} \end{bmatrix}.$$

Since points A', B', C' are non-colinear, \mathbf{T} is always invertible. The coefficients \mathbf{c} can be solved with $\mathbf{c} = \mathbf{T}^{-1}\mathbf{b}$. Let $|D'E'|$ be the dimension of an arbitrary object of interest on the plane $A'B'C'$. With their projections on the camera plane $D(x_D, y_D, z_0)$, and $E(x_E, y_E, z_0)$, we can obtain the line equations of OD , OE using their two-point forms. The coordinates of $D'(x'_D, y'_D, z'_D)$ and $E'(x'_E, y'_E, z'_E)$ are solutions of the equations, which represent the intersection of the lines OD , OE with the object plane $A'B'C'$

$$\begin{aligned} \mathbf{d} &= \mathbf{P}_D \mathbf{T}^{-1} \mathbf{b} \\ \mathbf{e} &= \mathbf{P}_E \mathbf{T}^{-1} \mathbf{b} \end{aligned} \quad (4.6)$$

$$\text{where } \mathbf{d} = \left[\frac{1}{x_D}, \frac{1}{y_D}, \frac{1}{z_D} \right]^T, \mathbf{e} = \left[\frac{1}{x_E}, \frac{1}{y_E}, \frac{1}{z_E} \right]^T, \mathbf{P}_D = \begin{bmatrix} 1 & \frac{y_D}{x_D} & \frac{z_0}{x_D} \\ \frac{x_D}{y_D} & 1 & \frac{z_0}{y_D} \\ \frac{x_D}{z_0} & \frac{y_D}{z_0} & 1 \end{bmatrix}, \text{ and}$$

$$\mathbf{P}_E = \begin{bmatrix} 1 & \frac{y_E}{x_E} & \frac{z_0}{x_E} \\ \frac{x_E}{y_E} & 1 & \frac{z_0}{y_E} \\ \frac{x_E}{z_0} & \frac{y_E}{z_0} & 1 \end{bmatrix}.$$

Once the coordinates of each end of a segment $D'E'$ is known, it is straightforward to calculate the magnitude of $D'E'$ in the direction \overrightarrow{DE} .

4.3 MODIFIED LASER BEAM MODEL AND CORRESPONDING ALGORITHM

4.3.1 Estimation of Approximate Intersections' Coordinates and Planar Patch Equation

In the implementation of 3D plane reconstruction, because the laser beams and projection rays are all estimated from the information extracted from images, the line equations are usually distorted by measurement noise and processing noise. As a result, two estimated skew lines never intersect at a certain point under real experimental conditions. However, with well controlled calibration and feature detection, the minimum distance between two skew lines is very small and an approximated intersection point can be calculated using a gradient descent searching method.

First, we measure the spatial line equation through a calibration step. There are several ways to represent the equations of OA , OB , and OC . We choose to use the two-point form equation in this method. Therefore we need to measure: 1) Points A' , B' , and C' with their pixel coordinates under the image coordinate system; 2) camera intrinsic matrix; 3) camera optical center position O ; 4) and the coordinates of A , B , and C on the camera plane. Details of the calibration process and coordinates transformation will be provided in Chapter 5.

It is an algebraic problem to solve the intersection of two skew lines. Since all the six line equations have been obtained from above analysis, we are able to estimate the intersections of Line I and OA , Line II and OB , and Line III and OC . The six line equations are:

$$\begin{aligned} \text{Line I : } \begin{cases} x = x_1 + a_1t \\ y = y_1 + b_1t \\ z = z_1 + c_1t \end{cases} & \quad \text{Line II : } \begin{cases} x = x_2 + a_2t \\ y = y_2 + b_2t \\ z = z_2 + c_2t \end{cases} & \quad \text{Line III : } \begin{cases} x = x_3 + a_3t \\ y = y_3 + b_3t \\ z = z_3 + c_3t \end{cases} \\ \\ \text{OA : } \begin{cases} x = x_A + a_At \\ y = y_A + b_At \\ z = z_A + c_At \end{cases} & \quad \text{OB : } \begin{cases} x = x_B + a_Bt \\ y = y_B + b_Bt \\ z = z_B + c_Bt \end{cases} & \quad \text{OC : } \begin{cases} x = x_C + a_Ct \\ y = y_C + b_Ct \\ z = z_C + c_Ct \end{cases} \end{aligned}$$

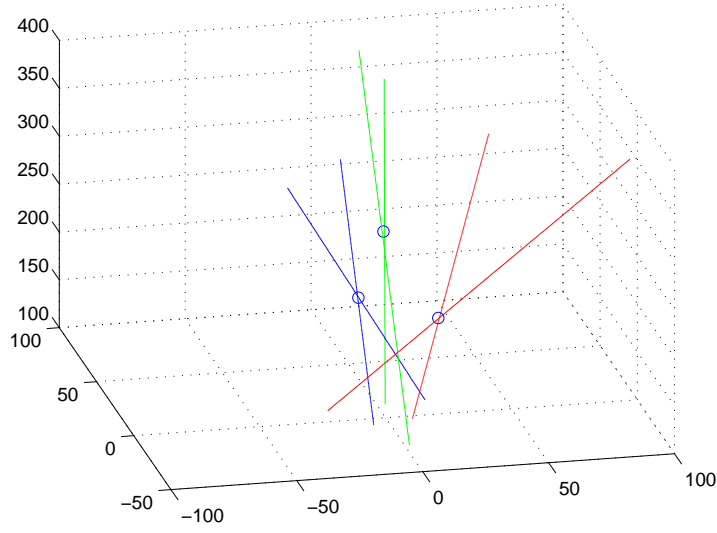


Figure 17: Six skew lines and estimate intersections

Figure 17 shows the six spatial lines and the approximated intersections. A search method as described in Table 12 (see Appendix) is adopted to search for two points on the two skew lines that have the shortest distance between them, and the point in the middle of them is used as the approximated intersection.

With the coordinates of three approximated intersections $A'(x'_A, y'_A, z'_A)$, $B'(x'_B, y'_B, z'_B)$, $C'(x'_C, y'_C, z'_C)$, we are able to determine the equation of the object plane, which contains the three intersections, as

$$\begin{vmatrix} x - x'_A & y - y'_A & z - z'_A \\ x - x'_B & y - y'_B & z - z'_B \\ x - x'_C & y - y'_C & z - z'_C \end{vmatrix} = 0 \quad (4.7)$$

4.4 ASSUMPTIONS AND GEOMETRIC RELATION AMONG IMAGE, CAMERA PLANE, LED BEAM AND OBJECT PLANE

In our second design, an LED spotlight is used as a reference light. A spotlight, which forms a conic region, is projected from the LED onto the object plane, as shown in Figure 18. The spot pattern, represented by a feature ellipse, is the intersection of the object plane and the lightening cone. The object plane intersects the optical axis (*center line*) of the LED at point P' . The distance between the camera optical center O and the LED optical origin O' is L . We assume that the optical axis of the camera and the center line of the LED cone are parallel, which can be controlled when we setup the device. Let the plane determined by the two parallel axes, the z axis and LED optical axis, to be Π . We assume that the pre-measured falloff angle of the LED cone is β and the dihedral angle between object plane and a vertical plane is θ , which is produced by rotating object plane along y axis only. One of the semi-axes, which is located in the plane Π , is r .

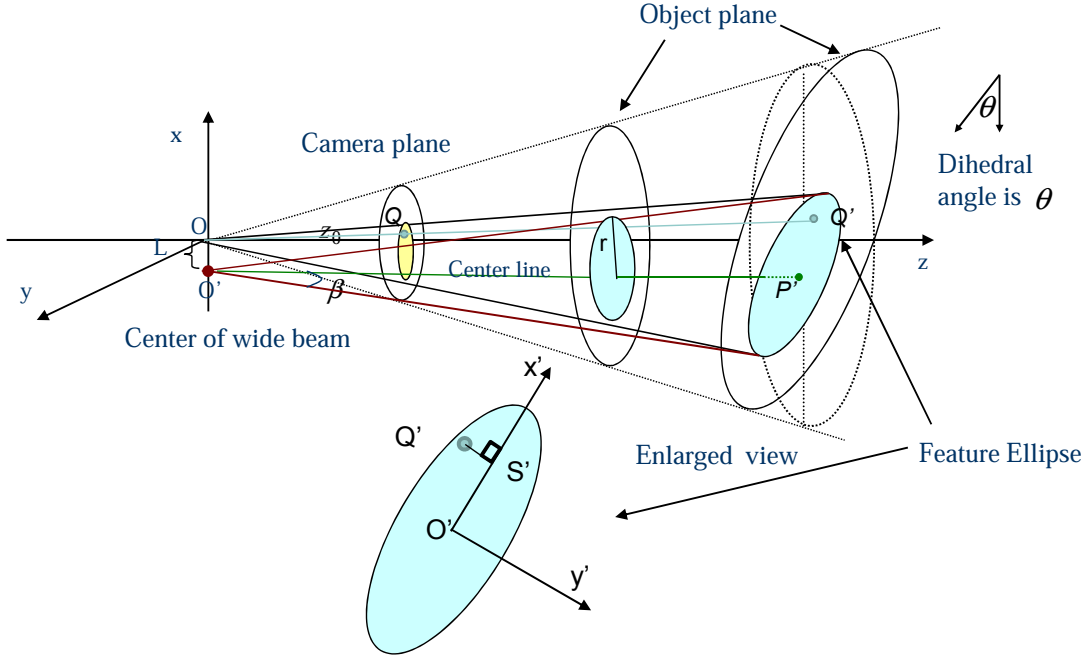


Figure 18: Geometrical relationships among image, camera plane, LED spotlight and object plane

4.5 LED BEAM SYSTEM MODEL AND CORRESPONDING ALGORITHM

The equation of the feature ellipse on the object plane can be derived from the following equation [106]:

$$(\cos^2 \beta - \sin^2 \theta) \cdot x'^2 + \cos^2 \beta \cdot y'^2 + r \cdot \sin 2\beta \cdot \sin \theta \cdot x' - r^2 \cos^2 \beta = 0 \quad (4.8)$$

.

If we take an arbitrary point Q' on the 2D feature ellipse with coordinate (x', y') , the projection of Q' on the camera plane is Q with coordinates (x, y) and S' is the perpendicular foot of the point Q' to the x axis on ellipse plane. With the knowledge of camera focal length z_0 and the distance L , and denoting that the distance between point S' and the camera plane is l , we can easily obtain the relation between the coordinates of object Q' and its image Q in their own coordinate system as

$$\begin{cases} \frac{x}{L+x' \cdot \cos \theta} = \frac{z_0}{z_0+l-x' \cdot \sin \theta} \\ \frac{y}{y'} = \frac{z_0}{z_0+l-x' \cdot \sin \theta} \end{cases} \quad (4.9)$$

x', y' can be solved from Eq. (4.9),

$$\begin{cases} x' = \frac{(z_0+l) \cdot x - z_0 \cdot L}{x \cdot \sin \theta + z_0 \cdot \cos \theta} \\ y' = \frac{L \cdot \sin \theta + (z_0+l) \cdot \cos \theta}{x \cdot \sin \theta + z_0 \cdot \cos \theta} y. \end{cases} \quad (4.10)$$

The radius r can be calculated by

$$r = (z_0 + l) \cdot \tan \beta. \quad (4.11)$$

Substituting Eqs. 4.10 and 4.11 into Eq. 4.8, we have

$$x^2 + \left(1 + \frac{L}{z_0 + l} \tan \theta\right)^2 \cdot y^2 - 2 \frac{z_0 \cdot L}{z_0 + l} \cdot x + \frac{z_0^2 \cdot L^2}{(z_0 + l)^2} - f^2 \cdot \tan^2 \beta \cdot \left(1 + \frac{L}{z_0 + l} \tan \theta\right)^2 = 0. \quad (4.12)$$

Eq. 4.12 can be rewritten into a standard ellipse equation :

$$\frac{(x - x_0)^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (4.13)$$

where $x_0 = \frac{z_0 \cdot L}{z_0 + l}$, $a = z_0 \cdot \tan \beta \cdot (1 + \frac{L}{z_0 + l} \tan \theta)$, and $b = z_0 \cdot \tan \beta$.

Therefore, the image of the light pattern projected on a tilted plane is also an ellipse. To eliminate the measurement of falloff angle β , we define a dissimilarity measure as

$$dissimilarity = \frac{a}{b} - 1 = \frac{L}{z_0 + l} \tan \theta. \quad (4.14)$$

The distance l can be estimated from the position of the ellipse center x_0 . The oblique angle θ of the object plane can be calculated from the dissimilarity index. For a given l , the greater the dissimilarity, the more elongated the ellipse, and hence the more tilted the object. After estimating l and θ , the coordinates of an arbitrary point in the object plane can be obtained from its image according to Eq. 4.10, and the equation of the planar patches is determined by the center and oblique angle of the ellipse. In this manner, the actual object size can be calculated.

The above algorithm can estimate an object's dimensions when the object plane is rotated along y axis. In fact, an object plane can be rotated along any combination of the three axes. However, the analytical expression of an arbitrary rotation is too complicated to derive. We analyze the rotation of an object plane separately with two independent parts: along y axis and along x axis, since the rotation along z axis will not affect our algorithm. The solution of a rotation along the x axis is similar with Eq. 4.10.

4.6 HEIGHT AND VOLUME ESTIMATION OF REGULARLY SHAPED OBJECTS

Using the algorithms described above, we are now able to estimate the 2D dimensions on the determined object plane from only one image. However, in many 3D reconstruction applications, the estimation of object volume is required. Fortunately, for certain regularly

shaped objects, prior knowledge of their profiles is available. We can use some clues from the object's shape and their profiles to estimate their height and volume.

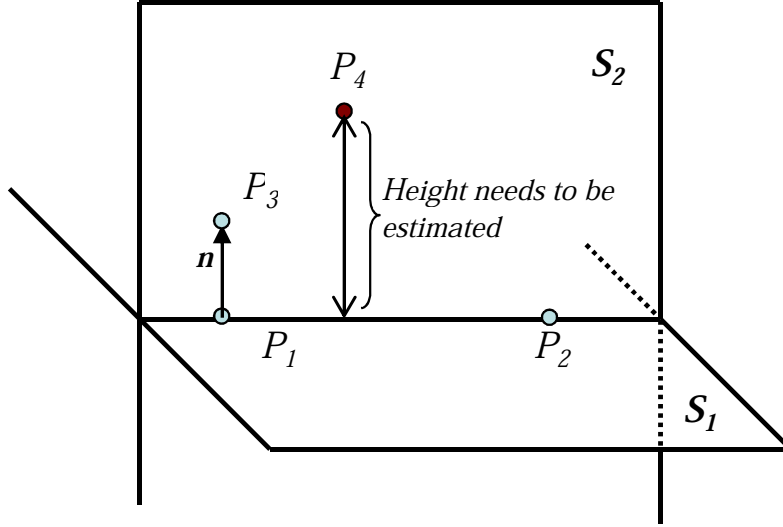


Figure 19: Height estimation on a perpendicular plane S_2 .

No information is available about the third dimension of the object besides what we can derive from the reference lights. However, we can utilize the shape of an object to find an equivalent height of the food, which can be estimated either manually or computationally. To simplify the description of the following derivation, we rename the object plane $A'B'C'$ as plane S_1 and the equation of S_1 is given by $c_1x + c_2y + c_3z = 1$. As shown in Figure 19, if there is a perpendicular plane S_2 in the image, we can easily find two points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ on the line, which is the intersection of the two planes. Since the normal vector $\mathbf{n}(c_1, c_2, c_3)$ of plane S_1 , which passes point P_1 , is on the plane S_2 , we can obtain another point $P_3(x_3, y_3, z_3) = P_1 + \mathbf{n}$ on the plane S_2 . With the coordinates of three points P_1 , P_2 , and P_3 , it is straightforward to calculate the equation of the plane S_2 using a three-point form as

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x - x_2 & y - y_2 & z - z_2 \\ x - x_3 & y - y_3 & z - z_3 \end{vmatrix} = 0 \quad (4.15)$$

Assume that the height we need to estimate is available and one of its ends is on the plane S_2 . We can reconstruct the metric coordinates of this end from the image using the line-plane intersection approach. For instance, let $P_4(x_4, y_4, z_4)$ be the other end of the height segment on plane S_2 . The dimension D of the height is just the distance from P_4 to plane S_1 ,

$$D = \frac{|\mathbf{n} - \mathbf{w}|}{|\mathbf{n}|} = \frac{|c_1(x - x_4) + c_2(y - y_4) + c_3(z - z_4)|}{\sqrt{c_1^2 + c_2^2 + c_3^2}} = \frac{|1 - c_1x_4 - c_2y_4 - c_3z_4|}{\sqrt{c_1^2 + c_2^2 + c_3^2}} \quad (4.16)$$

where $\mathbf{w} = - \begin{bmatrix} x - x_4 \\ y - y_4 \\ z - z_4 \end{bmatrix}$.

As long as we can find a perpendicular plane which contains the desired height of the object, we can use the above method to estimate the height based on the knowledge of determined object plane S_1 and clues of the object's shape from the image. However, for objects with irregular and complicated shapes, we have to tolerate a larger error in the estimation of the third dimension, or employ more sophisticated methods which are part of the future work. In this dissertation, we manually select equivalent heights for various food objects.

5.0 SYSTEM CALIBRATION

There are several different coordinate systems in our dimensional estimation model, including image, camera, world, and structured lights coordinate systems. We need to first transfer them into a uniform coordinate system and then apply the proposed methods within this system. Two system calibration processes are implemented: 1) Camera system calibration that connects a 2D image system using pixel unit with 3D real world system using metric unit; and 2) Structured lights system calibration which connects each component of the structured light with the camera system. In this chapter, we provide the details of our calibration processes.

5.1 CAMERA SYSTEM CALIBRATION

5.1.1 Checkerboard Approach

Camera calibration, often referred to as camera resectioning, is a way of examining an image, or a video, and deducing what the camera situation was at the time the image was captured. It was used primarily in robotic applications, but modern software applications make it quite easy to achieve, even for home use.

More specifically, camera calibration is the process of determining the internal camera geometric and optical characteristics (*intrinsic parameters*) and the 3D position and orientation of the camera coordinate system to a certain world coordinate system (*extrinsic parameters*). Several factors need to be considered in the calibration stage, such as the fact that camera pixels are not necessarily square, or that images are obtained by Analog-to-Digital card.

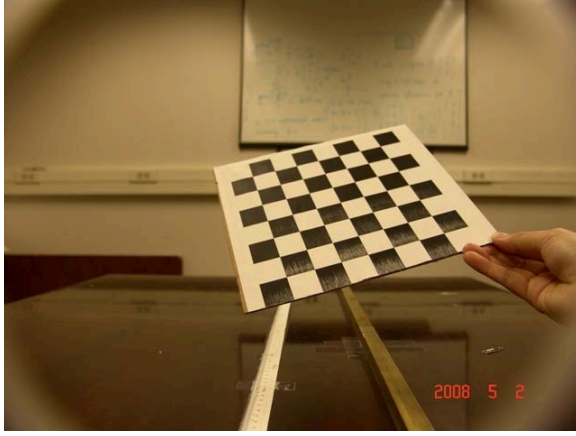
According to the dimensions of the calibration objects, we can classify the techniques into three categories: 3D reference object-based calibration [48, 51], 2D plane-based calibration [55, 56], and self-calibration [45, 57]. Other techniques also exist such as vanishing points [39, 40, 59] for orthogonal directions, and calibration from pure rotation [104, 105]. In our work, we used a 2D plane-based calibration approach. The classical 2D approaches used in early years entail solving for a large number of calibration parameters which required large scale nonlinear search [52]. The conventional way of avoiding this nonlinear search is to use the approaches similar to Direct Linear Transformation (DLT) [53, 54] that solves for a set of parameters with linear equations, ignoring the dependency between the parameters and the lens distortion. Tsai [51] proposed a well-known fast approach using a real constraint called *radial alignment constraint* and a two-stage technique to optimize the camera parameters. The solution is generally designed for mono-view calibration. In contrast to Tsai's method, Zhang proposed a flexible technique which requires at least two different orientations of an observed planar pattern to calibrate a free motion camera [55]. Without expensive calibration apparatus and elaborate setup, the requirement of calibrating a camera is further simplified. To calibrate our camera system, we used a modified approach based on Zhang and Tsai's method with a public Matlab toolbox from Jean-Yves Bouguet [92] in our camera calibration.

We use a checkerboard to calibrate the camera intrinsic parameters as shown in Fig 20. The *intrinsic matrix* \mathbf{K} combines a normalized coordinate system (focal length $f = 1$) and additional transformation parameters of the camera, which is shown in Eq. 2.8

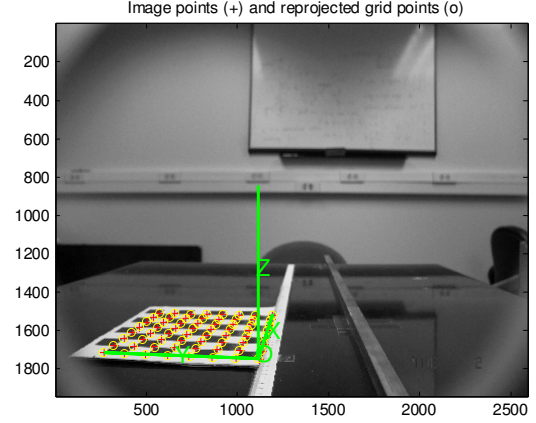
$$\mathbf{K} = \begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \quad (5.1)$$

The relation between normalized pixel coordinates in the image and normalized metric coordinates on the camera plane can be written as

$$\begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \quad (5.2)$$



(a) Intrinsic parameters calibration



(b) Extrinsic parameter calibration

Figure 20: Intrinsic and extrinsic parameter calibrations

where x_c and y_c are normalized by assuming the distance between camera optical center and camera plane is 1.

From the normalized camera plane coordinates in Eq. 5.2, the estimated coordinates of three points A , B , and C , which are the projections of all three laser spots on the camera plane (as shown in Figure 16) under the camera frame can be calculated using Eq. 5.3

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = f \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \quad (5.3)$$

Similarly, we transfer all the coordinates of other points, such as O' , A' , B' , and C' from the world frame to the camera frame by Eq. (2.1).

5.2 STRUCTURED LIGHTS CALIBRATION

Our method uses structured light to retrieve the distance as well as the orientation information of a planar object, therefore a precise calibration of the structured light is necessary. However, to manufacture and calibrate a precise equilateral triangle structure with three laser beams is very difficult. In fact, a perfect equilateral triangle structure is not necessary for our algorithm as long as we can measure the spatial equation of each laser beam individually. We designed a flexible and simple structured light calibration approach based on the Orthogonal Linear Regression (OLR) line-fitting method. This approach is also used to calibrate the LED spotlight.

5.2.1 Laser Beams Calibration

5.2.1.1 Orthogonal Linear Regression Method

In ordinary linear regression, the goal is to minimize the sum of the squared vertical distances between multi-dimensional data values and the corresponding values on the fitted line. In orthogonal regression, alternatively the goal is to minimize the orthogonal (perpendicular) distances from data points to the fitted line.

For example, the slope-intercept equation for a 2D line is:

$$Y = m \cdot X + b \quad (5.4)$$

where m is the slope and b is the intercept.

A line perpendicular to this line will have a slope of $-\frac{1}{m}$, and its equation will be

$$Y' = -\frac{X}{m} + b'. \quad (5.5)$$

If this line passes through some data point (X_0, Y_0) , its equation will be

$$Y' = -\frac{X}{m} + \left(\frac{X_0}{m} + Y_0\right). \quad (5.6)$$

The perpendicular line will intersect the fitted line at a point (X_i, Y_i) , where X_i and Y_i are defined by:

$$\begin{aligned} X_i &= (X_0 + m \cdot Y_0 - m \cdot b)/(m^2 + 1) \\ Y_i &= m \cdot X_i + b. \end{aligned} \tag{5.7}$$

Therefore the orthogonal distance from (X_0, Y_0) to the fitted line is the distance between (X_0, Y_0) and (X_i, Y_i) , which is computed by

$$d_i = \sqrt{(X_0 - X_i)^2 + (Y_0 - Y_i)^2}. \tag{5.8}$$

By minimizing the total orthogonal distances $\sum_i d_i$, we can fit 2D data to a line. The 3D spatial line fitting is quite similar by minimizing the three-dimensional total orthogonal distance

$$\sum_i \sqrt{(X_0 - X_i)^2 + (Y_0 - Y_i)^2 + (Z_0 - Z_i)^2} \tag{5.9}$$

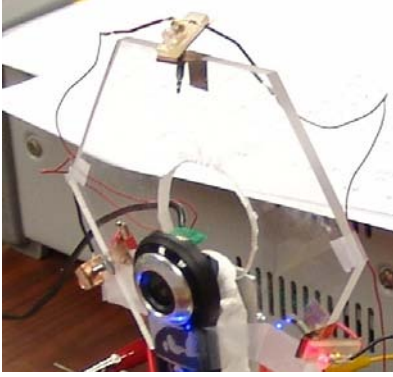
This fitting method allows our system to be more flexible since there is no constraint on the perfect alignment of the structured lights.

5.2.1.2 Measurement Of Spatial Line Equation of $O'A'$, $O'B'$ and $O'C'$

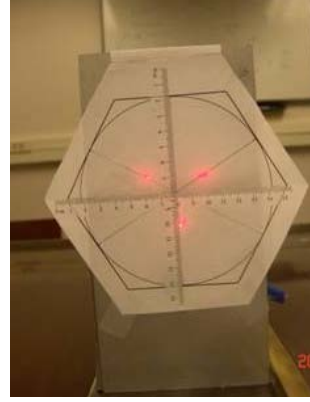
To calibrate the structured light with respect to the camera coordinate system, we propose an image processing approach to implement the sampling and feature localization. After that, the OLR algorithm is used to fit the samples into a spatial line.

First, we fix the camera and the structured lights at one testing panel (see Figure 21 (a)). The testing panel is placed parallel to the $x - y$ plane and its position is marked under a pre-defined world coordinate system. The measurement pattern (see Figure 21 (b)) is marked by rulers in both x and y directions and is aligned parallel to the testing panel well and fixed on a rail which allowed us to slide the measurement pattern along the z direction. When we turned on the laser or LED light, the highlighted spots were captured by the camera and shown in images like Figure 21 (b).

By sliding the measurement pattern to a sequence of controlled positions, we can read the x and y coordinates of the spot's center from the captured image sequence and read the corresponding coordinate on z direction from the marked rail (see Figure 22). We record the



(a) Testing board with laser modules and a camera



(b) View captured by the camera

Figure 21: Structured light calibration system setup

sampling measurements from multiple positions for each of the laser beams or LED spotlight and provide the resulting data to an OLR estimator to estimate the spatial line equations in a least-squared sense. Figure 23 shows the OLR fitting results for three laser beams in a 3D view and a 2D side view. The circles represent the sampled positions and the three colored solid lines are the approximated lines for each of the laser beams. As shown in Figure 23 (b), three laser beams do not intersect at one point because the three laser modules could not be centered at a single point in the real experimental instruments. This drawback can be eliminated if we use a single laser module with a diffraction grating film. We proposed a modified method in Sec. 4.3 to deal with this less than ideal structured light situation. As long as we can measure each beam's spatial equation, our proposed algorithms can estimate the dimensions from a single view.

Although the laser beams are narrow and well focused, they diverge as the distance increases. We use the center point of a laser highlighted spot to represent the intersection of a laser beam and the object. The method to extract the highlighted spot center point will be described in Chapter 6.

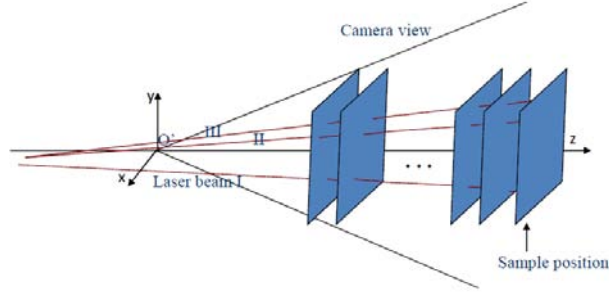
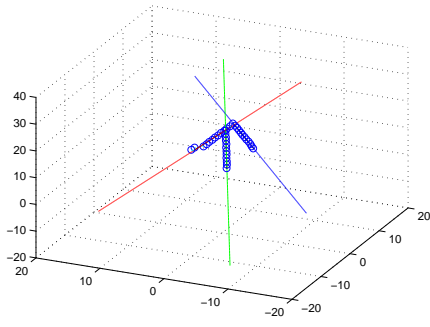


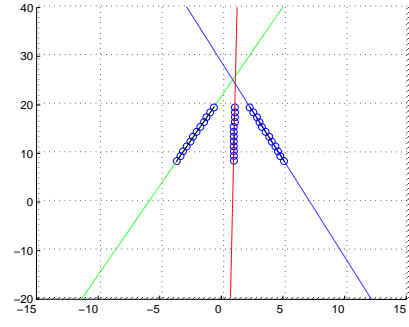
Figure 22: Sampling along the z direction

5.2.2 LED Beam Calibration

The calibration of the LED spotlight is similar to the calibration of the laser beams. To implement the algorithm which uses LED circular feature to estimate the position and orientation of an object plane, we need to know the center of the LED spot and the ratio between the major semi-axis and minor semi-axis. In the calibration process, we mount a calibrated camera and an LED diode on a testing panel which is parallel to the $x - y$ plane of the world coordinate system, and then move the measurement panel along the z direction. The experimental environment is shown in Figure 24 (a) and one snapshot of the measurement panel with a highlighted spot is shown in Figure 24 (b). To find the center of the LED spot from images, we need to detect the contour of the highlighted region and fit the segmented region into an ellipse. Figure 24 (c) shows the estimated contour of an LED highlighted spot in one of the measurement images. The center of the LED spot is the center of the detected ellipse. We will provide the details of the ellipse extraction in Chapter 6.

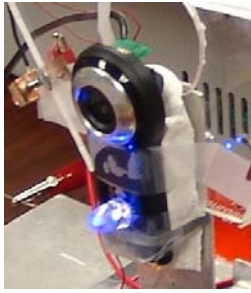


(a) 3D view of three spatial skew lines

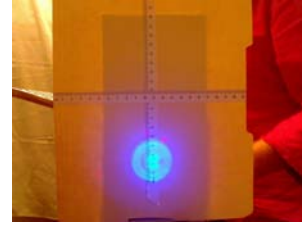


(b) 2D view of three spatial skew lines

Figure 23: Spatial lines fitting using Orthogonal Linear Regression



(a) The setup of laser and LED diode calibration



(b) A highlighted spot on a measurement panel



(c) LED pattern extraction and fitting

Figure 24: LED diode calibration experimental setup, measurement and fitting

6.0 CIRCULAR PATTERN EXTRACTION

In order to estimate the position and orientation of the projected structured lights, such as laser and LED spotlight, we must extract the structured light patterns from the background of the captured images and estimate their parameters. This feature extraction problem involves multiple tasks, such as segmentation, shape fitting, parameter estimation and localization.

Given a digitized image containing several objects, the feature extraction process consists of two major phases. The first phase is *image segmentation* in which the region of interest (ROI) is isolated from the rest of the scene. The second phase is *feature extraction* where the objects are measured and a set of features (usually comprising a *feature vector*) representing some significant characteristic of the objects is produced. Feature vector represents the necessary knowledge upon which subsequent classification decisions are based. This drastically reduced the amount of information compared to the original image or segments of the image. In this chapter, we will introduce our solutions to these problems based on particular characteristics of the structured lights that we used in our devices.

6.1 TECHNIQUES FOR CIRCULAR FEATURE PATTERN EXTRACTION

6.1.1 Pattern Segmentation

Thresholding is a particularly simple and effective technique to segment scenes containing solid objects resting upon a contrasting background [116]. Generally, the thresholding methods can be categorized into six groups [107]: histogram shape-based methods, clustering-

based methods, entropy-bases methods, object attribute-based methods, spatial methods using higher-order statistics, and local methods. There have been a number of survey papers on thresholding. Lee conducted a comparative analysis of five global thresholding methods and studied advanced useful criteria for thresholding performance evaluation [108]. Trier and Jain provided an extensive comparison basis (19 methods) in the context of character segmentation from complex image background [109]. Glasbey demonstrated the relationships and performance differences among 11 histogram-based algorithms based on an extensive statistical study [110]. Thresholding works well if the objects of interest have a nearly uniform gray level. After the thresholding process, the image is converted into a binary one representing the object and background classes. In this dissertation, we propose to use a single channel or a linear combination of two or three channels to map the original color image into a gray level image. Then, we use thresholding methods to segment the ROI.

Laser has "coherent light", which denotes a light source that produces (emits) light of in-step waves of identical frequency, phase, and polarization. The laser modules used in our device produce known monochromatic light. However, considering the reflection characteristics of different materials, color, and position and orientation of the reflection surface, the observed color and shape of a laser spotlight in images varies. We investigated the color properties of selected laser modules and LED diodes and proposed appropriate methods in Section 6.2 to segment their signatures from various image backgrounds. The output of this segmentation is a binary image containing 1 and 0 to represent the ROI and background or inverse.

6.1.2 Feature Vector Extraction

To extract the features after segmentation, without loss of generality, we assume that the laser module produces an isotropic diverging narrow monochromatic beam. We will assume that the reflecting spot is a circle in the captured image. A three-dimensional feature vector, $V_{laser} = \{x_0, y_0, r\}$, is used to represent a circle in the feature space in our problem. Here, x_0 and y_0 are the coordinates of the center of circle in the image, and r is the radius. For the LED diode case, we use a right cone with a known falloff angle to represent the lightening

region. The intersection of this cone and the object plane has been shown in Chapter 4 to be an ellipse. Therefore, a five-dimensional feature vector, $V_{LED} = \{x_0, y_0, a, b, \theta\}$, is used to represent an ellipse. Here, x_0 and y_0 are the coordinates of the ellipse center, a and b are the major and minor axes, and θ is the orientation of the ellipse.

To estimate the corresponding feature vectors of a circle or an ellipse, we first perform an edge detection on the binary image which is the output of the segmentation phase. Normally, we will have discontinuous edge segments after segmentation. To obtain a reasonable feature vector, we utilize an ellipse fitting method proposed by Taubin [111] to estimate the five parameters with a Matlab package developed by Nikolai Chernov [112]. Taubin developed an approximate distance, which is a first-order approximation of the real distance, from a point to a curve or surface, and turned the problem of fitting curves and surfaces into the minimization of the approximate mean square distance. They showed that this nonlinear least-squared problem can be reduced to a generalized eigenvector fit for certain families of nonsingular curves and surfaces and presented a variable-order segmentation algorithm based on the algorithm.

After fitting the detected edge into an ellipse, we know the position of the ellipse center and the ellipse's orientation. For the laser beams design, the circular feature spots became ellipses with the same major and minor axes. The centers of three ellipses are used to determine an object plane. For the LED spotlight design, the plane determined by the ellipse is the plane of object.

There may be multiple objects, especially an elliptic object, with similar color and shape as the structured lights in the image. Therefore, more information is needed to better recognize the structured light, such as the possible position range and the size of the features. Though we currently use a human-supervision approach to classify the desired features from other objects, we will pursue the auto-classification of the feature patterns in our future work.

6.2 COLOR-BASED THRESHOLDING SEGMENTATION AND PATTERN EXTRACTION

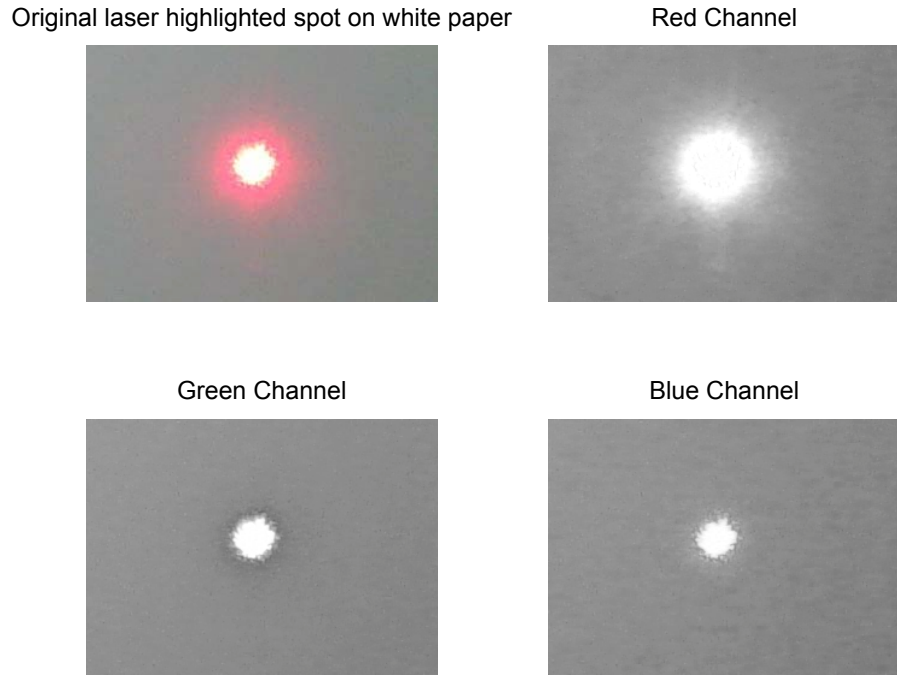
6.2.1 Laser Pattern Extraction

In the first device prototype design, we used three narrow-beam red laser modules to construct a triangle-shape structured light. Since the laser beam diverges slightly as the distance increases, usually we can detect a circular or an elliptical red spot in the observed image. The center point of a highlighted spot is used to represent the intersection of the laser beam and the object surface. The intensity of these spots varies in different images depending on the illumination conditions and the reflection property of the object surface. Intuitively, we first split the color image into three channels: red(R), green(G), and blue(B) channels. Figure 25 (a) shows the original color image and the split R-channel, G-channel, and B-channel of a red laser spot reflected by a white paper. We can distinguish the difference between R-channel and the other channels within the region of the laser spot. In the region around the center of the spot, intensity is saturated. Therefore all channels have high values. The R-channel shows a larger high intensity region because the laser emits red light. Figure 26 shows the 3D shapes of the pattern intensity in the region of laser spot in three channels with a uniform background (white paper). However, this difference changes as the background object changes. For instance, when the red laser beams are reflected by a slice of wheat bread or a slice of yellow cheese, the absolute intensity in the spot region at each channel has considerable difference. Since no uniform threshold can be set for different backgrounds, a fixed intensity value is not a robust threshold to separate the ROI from the background.

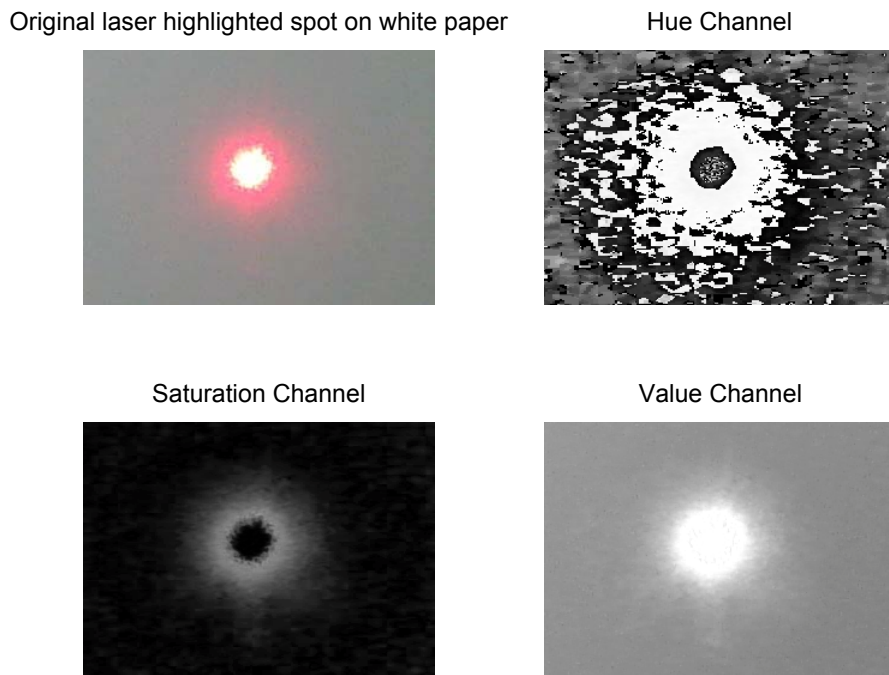
HSV(Hue, Saturation, Value) or related models, such as HSL (Hue, Saturation, Lightness), or HSI (Hue, Saturation, Intensity), are often used in computer vision and image analysis for feature detection or image segmentation. The applications of such tools include object detection for robot vision, object recognition, text or license plates detection, content-based image retrieval, and analysis of medical images [113]. The motivation for using the HSV model is that RGB color models does not define color relationships the same way the human eye does and some researchers think that users are better at describing colors in

terms of HSL than RGB coordinates [114]. Since our structured light has particular known colors (red and blue-white), we split the color images into HSV channels and apply the segmentation process to them. Figure 25 (b) shows the splitting results of a laser spot on a piece of white paper. We can see the ROI has obviously higher intensity compared with the backgrounds at the V-channel.

We want to maximize knowledge of the monochromatic and intensity superposition of the spot patterns. Upon further investigation of the difference between R-, G-, B-channels, we found the *Difference-Image* between the red and green channels is also a possible candidate for applying a global thresholding method. We used two different thresholding methods, a popular clustering-based thresholding method proposed by Otsu [115] as well as a histogram-based thresholding method, on different reflective surfaces with three candidate images: Red-channel, Difference-Image, and Value-channel. Otsu’s method minimizes the weighted sum of within-class variances of the foreground and background pixels to establish an optimum threshold. The goal then is to select the threshold that minimizes the combined spread. From the segmentation results, we find that for a clean background, good outer edges can be obtained from all of the three candidate images with Otsu’s threshold and histogram-based method. Figure 30 shows the segmentation using Otsu’s method on three gray-level images. The ellipse fitting process is then applied to these segmentation results. The solid line represents the fitting result and the ellipse center is marked by a star. The Difference-Image has a hole in the middle because of the saturation in both R- and G-channels of the laser spot. With a complicated background, the histogram-based method performs better than the Otsu’s method, likely because two clusters are not enough to separate the features from various background contents. With the histogram-based thresholding method, we slightly adjusted the threshold values based on experimental performances. For the R-channel and V-channel images, we used a value between 0.75 to 0.80. For the Difference-Image, we used a value between 0.18 to 0.3. Figure 27, 28, and 29 show the segmentation results using both thresholding methods. The results obtained from a number of experiments have shown the validity of the presented method for satisfactory practical use.



(a)



(b)

Figure 25: Intensity comparison of channel splitting : (a) A laser spot on a white paper background in Red, Green, and Blue channels. (b) A laser spot on a white paper background in Hue, Saturation, and Value channels.

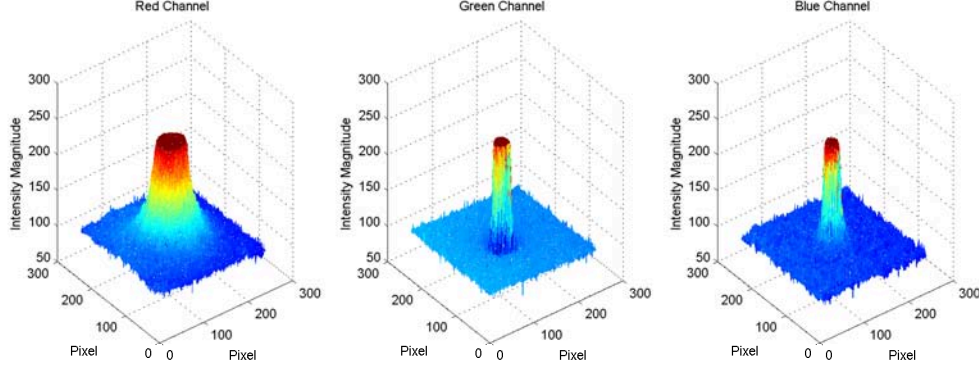
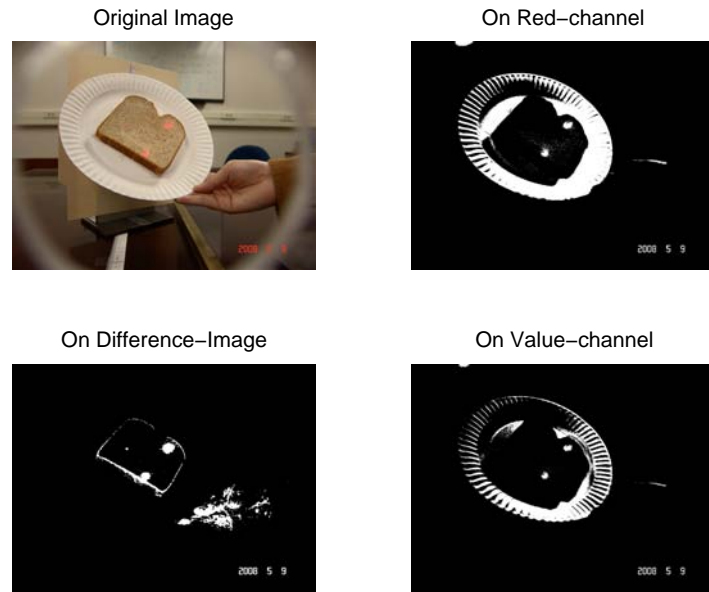


Figure 26: 3D shapes of intensities in Red, Green, and Blue channels.

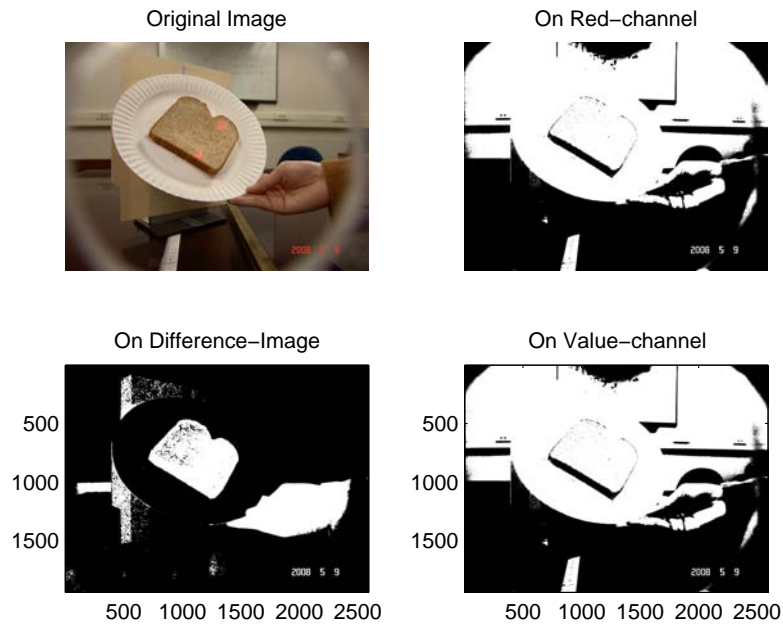
6.2.2 LED Pattern Extraction

Apart from the red laser beams, we used a blue-white light LED diode in the LED structured light design. The monochromatic advantage of the laser beams device is no longer available with this design. However, we found that the ROI has relatively better contrast in the R-, G-, H-, and S-channels (see Figure 31 (a) and (b)). We believe that is because the LED light superposes extra illumination on the object surface and the reflections of these regions have a naturally higher intensity. For this reason, we select the R-, G-, H-, and S- channels as the inputs for the segmentation phase. A comparison is made using histogram-based and clustering-based segmentation methods on different channels with other objects in the image. The combination of the G-channel and histogram-based thresholding method has shown the best segmentation performance. From the segmentation results shown in Figures 32 and 33, the G-channel image is shown to be the best candidate for segmentation, and the histogram-based method has shown better performance than Otsu's method. Therefore, we used the combination of G-channel and histogram-based thresholding method in the LED feature extraction phase.

After obtaining a successful segmentation, we apply the ellipse fitting method to estimate the feature vector on the desired objects. The fitting results are shown in Fig 34. The solid lines are the fitted ellipses and the stars are the centers of ellipses.

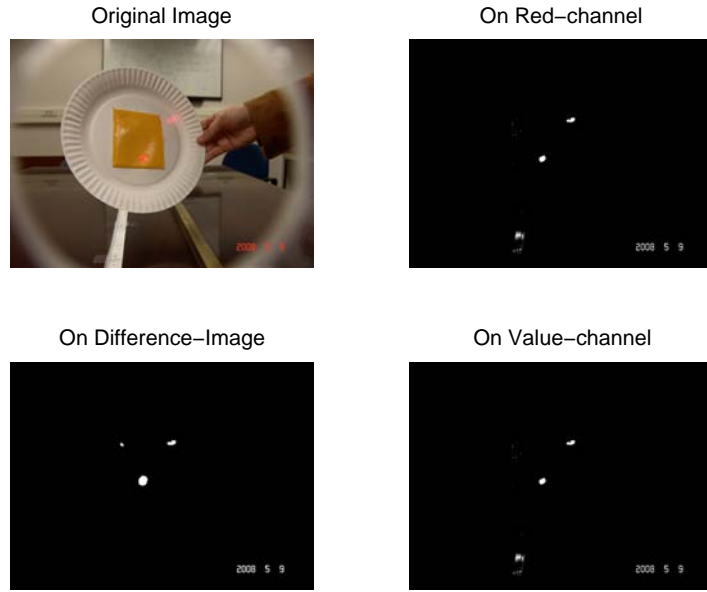


(a)

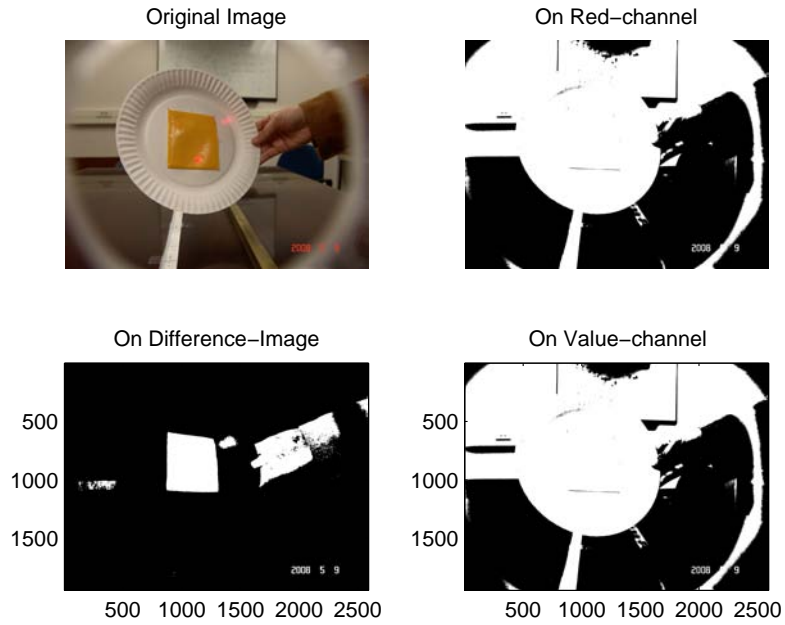


(b)

Figure 27: Performance comparison on wheat bread: (a) Histogram-based method; (b) Clustering-based method.

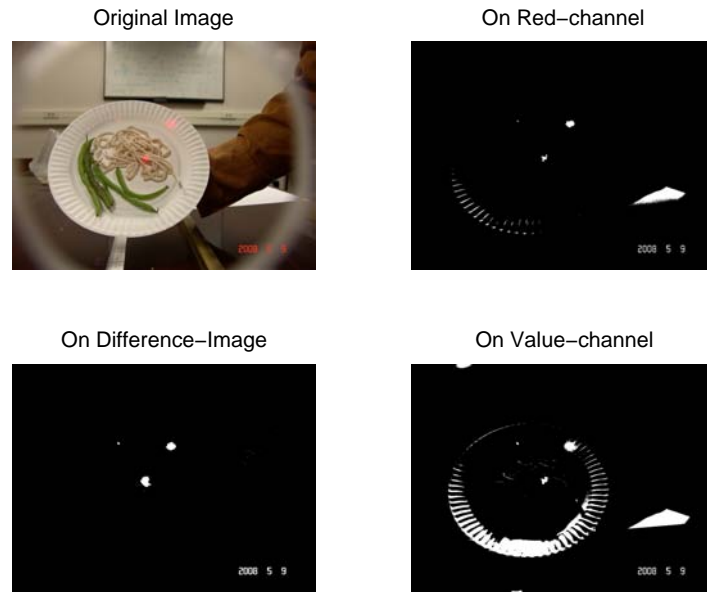


(a)

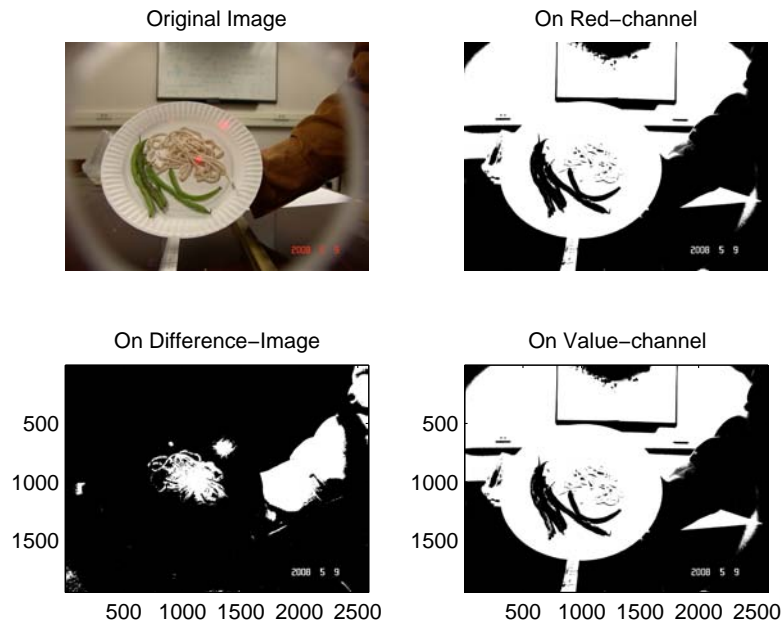


(b)

Figure 28: Performance comparison on yellow cheese: (a) Histogram-based method; (b) Clustering-based method.



(a)



(b)

Figure 29: Performance comparison on a plate of noodle and beans: (a) Histogram-based method; (b) Clustering-based method.

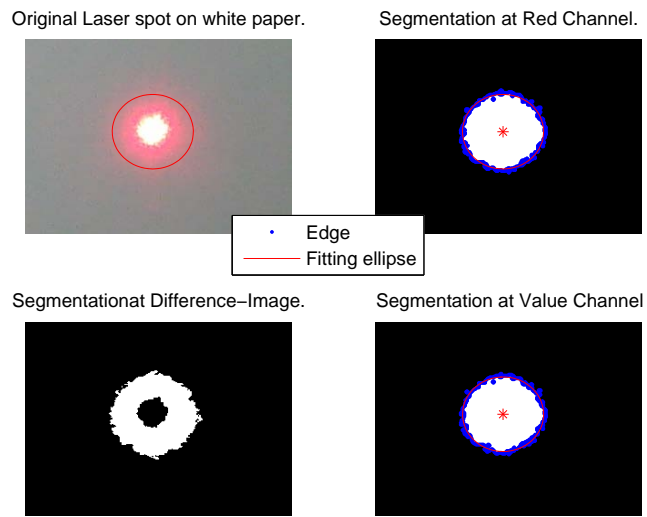
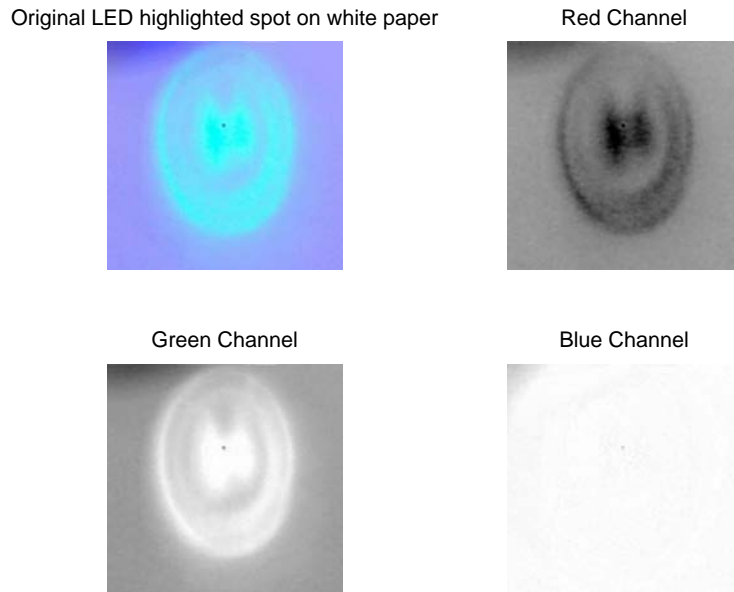
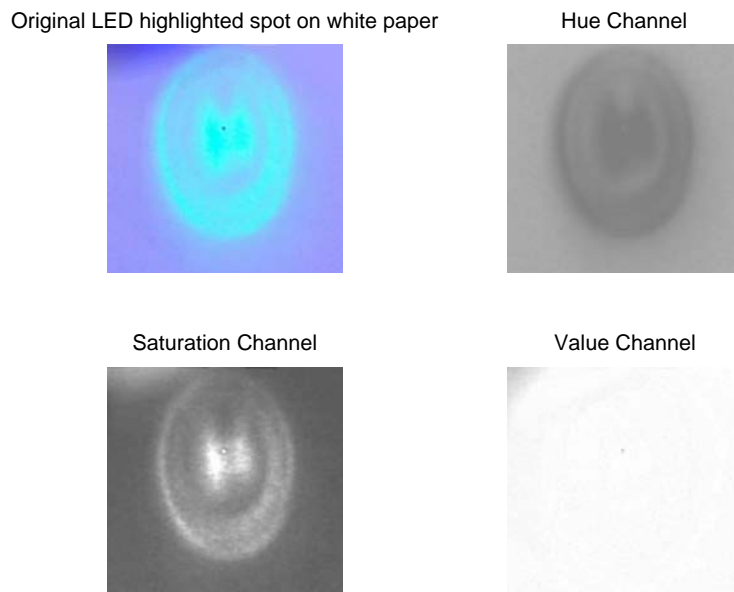


Figure 30: Segmentation on Red channel, Difference-Image, and Value channel and automatic ellipse-fitting are shown on the Red channel and Value channel.

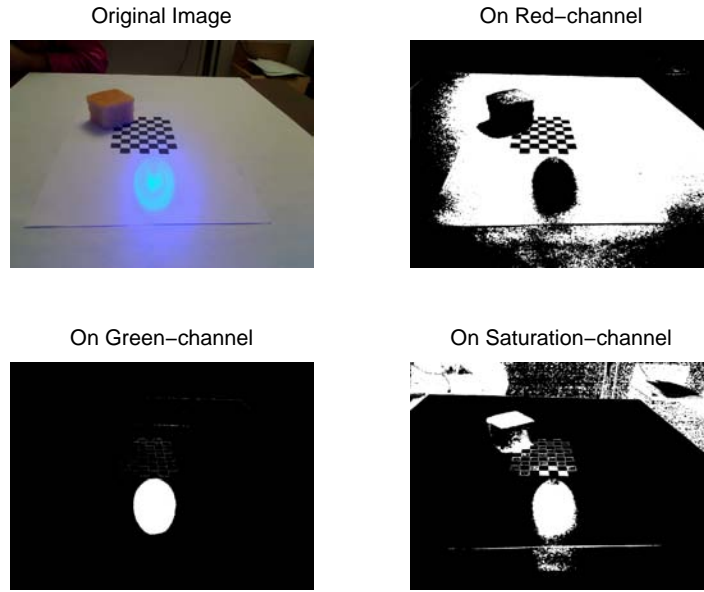


(a)

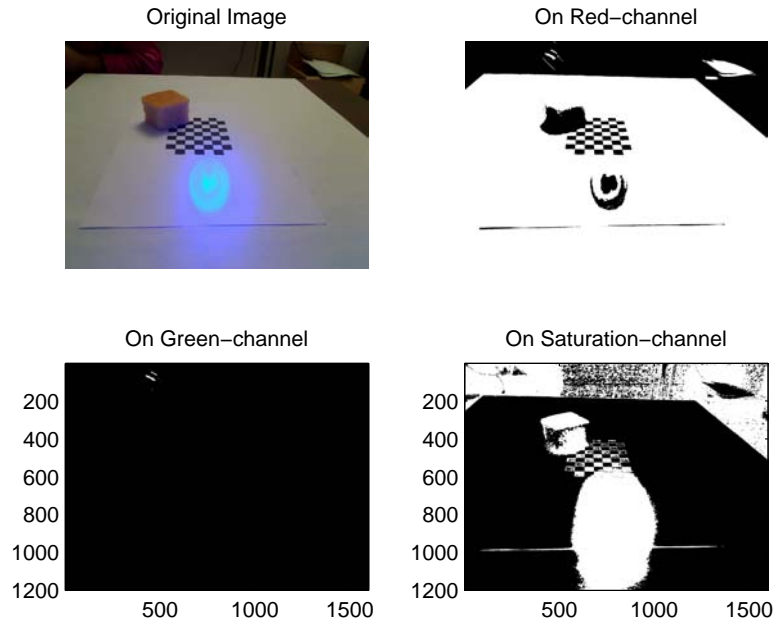


(b)

Figure 31: Intensity comparison of channel splitting : (a) An LED spot on a white paper background in Red, Green, and Blue channels. (b) An LED spot on a white paper background in Hue, Saturation, and Value channels.

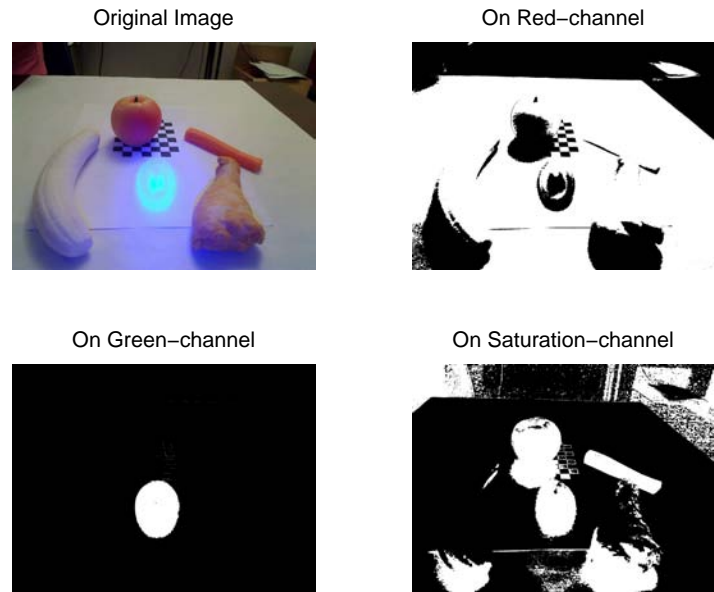


(a)

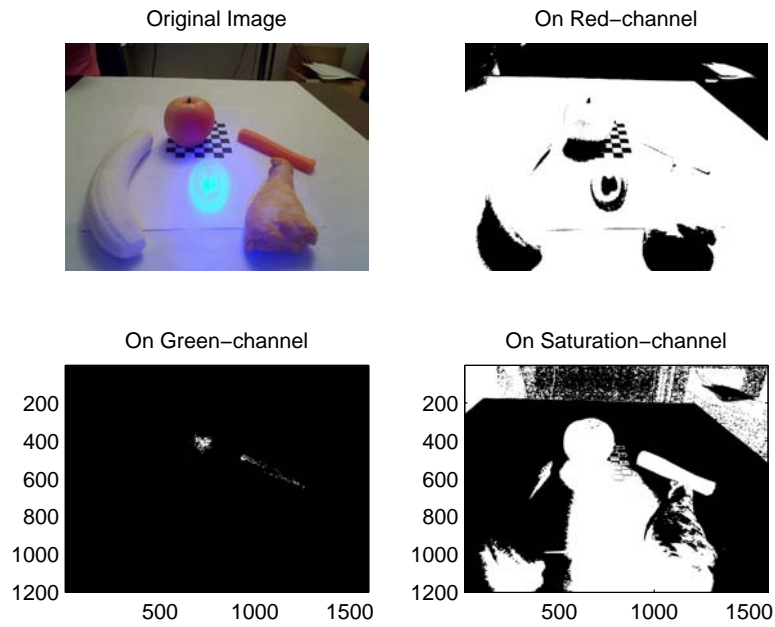


(b)

Figure 32: Performance comparison: (a) Histogram-based method; (b) Clustering-based method.



(a)



(b)

Figure 33: Performance comparison: (a) Histogram-based method; (b) Clustering-based method.

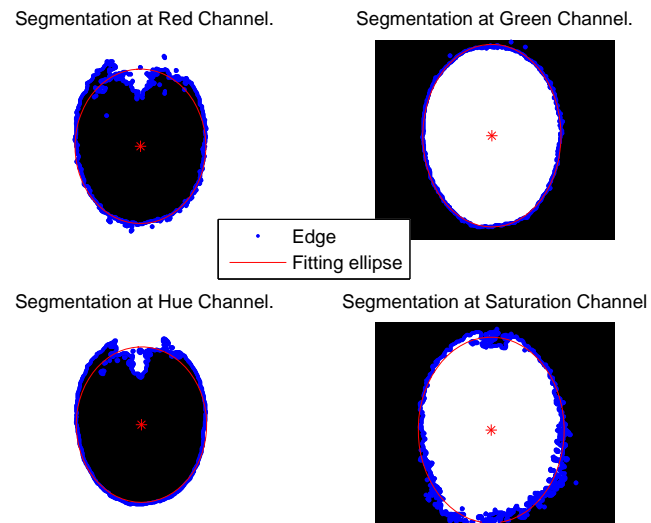


Figure 34: Segmentation and automatic ellipse fitting on Red, Green, Hue and Saturation channels.

7.0 EXPERIMENTAL RESULTS

We perform a number of experiments to verify the performance of the proposed prototypes, the laser beams structured light prototype, and the LED structured light prototype. Both artificial objects and real objects are used to test the performance of the proposed 3D dimension estimation algorithms. Two imaging instruments, a high resolution digital camera and a median resolution miniature camera were both tested to evaluate the system performance and robustness with different inputs. The experimental setup, estimation results, and further discussion are presented in this chapter.

7.1 EVALUATION OF LASER BEAMS DESIGN

In order to obtain a good segmentation of the laser highlighted spots, we chose to use three identical laser modules as reference lights because they had similar illumination intensities. The laser modules were mounted on a testing panel. The details have been described in Section 3.2.

7.1.1 Experiment with High Resolution Digital Camera

First, we performed an experiment with a high resolution digital camera as the image acquisition device (a Sony DSC-F828 with resolution 2592×1944 pixels). The instrument and system setup is shown in Figure 35. Several trials were conducted to test our laser beams geometric algorithm with both artificial objects and real food objects. The three laser modules were connected in series and driven by a voltage/current source. The operational voltage

and current of each laser diode was 2.1V and 20mA, respectively, and the operational power was about 40mW.



Figure 35: Prototype with laser beams structured light and high resolution digital camera used in our experiment.

The spatial equations of all three laser beams were estimated using the method discussed in section 5.2.1 as

$$\begin{aligned} \frac{x+30.5513}{-0.2125} &= \frac{y+29.3961}{-0.173} = \frac{z-438.8792}{0.9619} \\ \frac{x-68.7446}{0.341} &= \frac{y+25.6228}{-0.1445} = \frac{z-435.8595}{0.9289} \\ \frac{x-14.5248}{0.0389} &= \frac{y-66.6045}{0.3387} = \frac{z-437.9242}{0.9997} \end{aligned} \quad (7.1)$$

The camera calibration was performed using a checkerboard and a Matlab[®] camera calibration toolbox by Bouguet [92]. The measured intrinsic matrix was

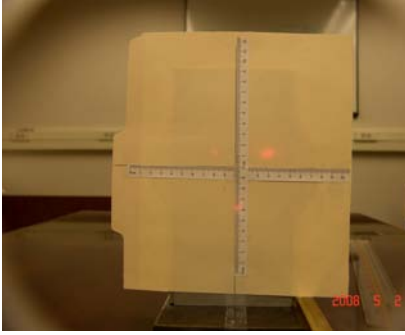
$$\mathbf{K} = \begin{bmatrix} 2203.1 & 0 & 1308.2 \\ 0 & 2216.2 & 973 \\ 0 & 0 & 1 \end{bmatrix}$$

The elements in \mathbf{K} were in the pixel units. The camera pitch size was given in manufacturer's specification sheet as $2.7 \mu\text{m}$. The measured fixed focal length was 5.95mm. The measured rotation matrix and translation vectors (in millimeters) were

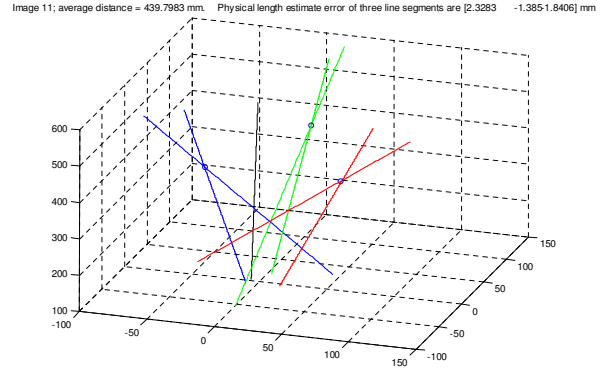
$$\mathbf{R} = \begin{bmatrix} 0.030564 & -0.999531 & 0.002113 \\ -0.004408 & -0.002248 & -0.999988 \\ 0.999523 & 0.030554 & -0.004475 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} -34.742984 \\ 138.076016 \\ 387.721180 \end{bmatrix}$$

Given a single image, we first extracted the feature vector, and then reconstructed the projection rays of the spots. For example, we estimated three highlighted spots' coordinates reflected from a paper board as shown in Figure 36 (a) based on the perspective projection theory. From their pixel coordinates in the image and the estimated parameters \mathbf{K} , \mathbf{R} , and \mathbf{T} , we obtained equations of the three projecting rays:

$$\begin{aligned} \frac{x}{-0.4262} &= \frac{y}{-0.36383} = \frac{z+5.95}{5.95} \\ \frac{x}{0.93561} &= \frac{y}{-0.31349} = \frac{z+5.95}{5.95} \\ \frac{x}{0.1897} &= \frac{y}{0.95254} = \frac{z+5.95}{5.95} \end{aligned} \quad (7.2)$$



(a) laser beams highlighted spots on a paper board



(b) Intersection estimation

Figure 36: Distance and dimension estimation on a perpendicular plane

In this experimental setting, the paper board with attached rulers was perpendicular to the table and parallel to the camera. Therefore, the three highlights had the same “ z ” distance to the camera origin. The measured distance between the paper board and the origin as shown in Figure 15 was 440 mm. The estimated “ z ” distances of the three points were 442.13 mm, 438.41 mm, and 437.96 mm, respectively, with an average of 439.7983 mm. Figure 36 (b) shows the three approximated intersection points. The solid line in the middle is the camera optical axis. With the coordinates of three intersection points, we estimated the distance between object and camera, as well as dimensions of arbitrary object on the board.

In this dissertation, we define two values to compare the estimation performance: *Absolute Error* and *Relative Error* as

Absolute Error = $|X - \hat{X}|$, and

Relative Error = $\frac{|X - \hat{X}|}{X} \times 100\%$. where X is the true dimension and \hat{X} is the estimated dimension using our algorithm.

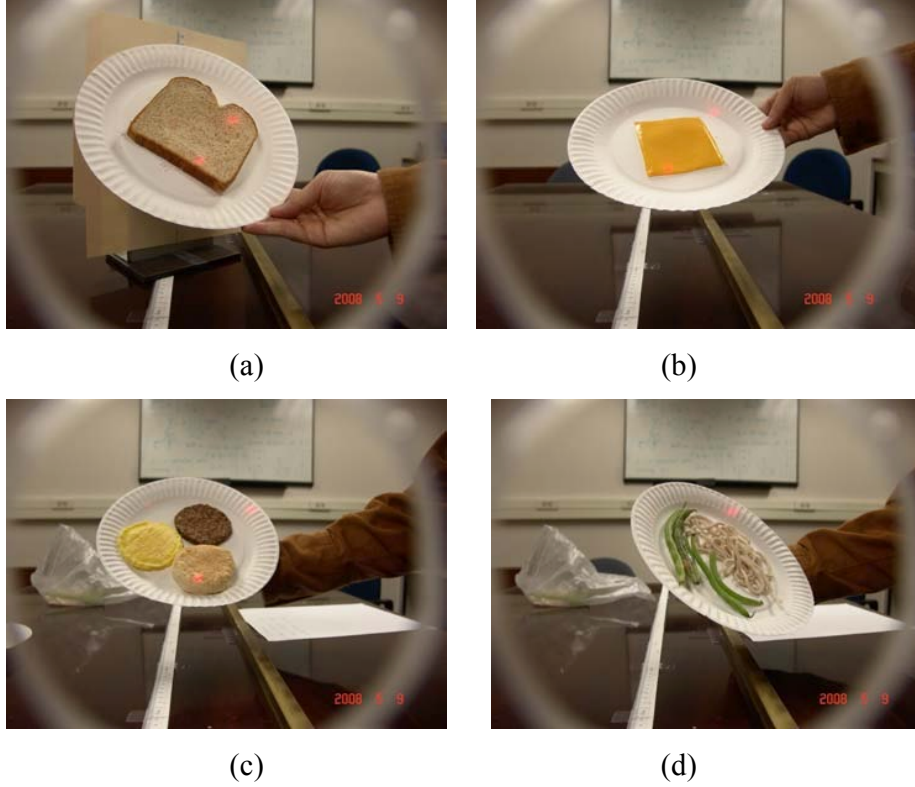


Figure 37: Real food samples for the dimension estimation experiments.

Various artificial objects and real foods were placed on a white paper plate (Figure 37) at random positions. The statistical estimation results of distance and physical dimension are listed in Table 1 and Table 2. The relative estimation errors of our method are less than 10% of the true dimensions.

Table 1: Distance and physical length estimation with artificial objects

	Distance to a vertical plane	Dimension on a vertical plane	Dimension on a rotated plane
Number of Trials	11	12	24
Absolute mean of error (mm)	2.0	3.3	3.7
Mean and STD of relative Error(%)	0.44 ± 0.76	2.1 ± 1.7	2.44 ± 2.31

Table 2: Real food physical length estimation at arbitrary position using laser structured light.

	Bread	Cheese slice	Round steak	Noodle
Number of Trials	13	15	12	14
True Dimension (mm)	107	76	80	130
Average Estimated Dimension (mm)	109.3	81.34	81.4	135.2
Mean of error (mm)	2.3	5.4	1.4	5.2
Mean and STD of relative Error(%)	2.1 ± 4.7	7.1 ± 5.4	1.8 ± 1.0	4.0 ± 5.0

7.1.2 Experiment with Miniature Camera

Secondly, we performed experiments with a median resolution miniature camera as the image acquisition device (Logitech web-camera with resolution 1600×1200 pixels). The prototype and system setup were shown in Figure 38. Several trials are conducted to test the laser beams geometric algorithm with a set of real food objects.

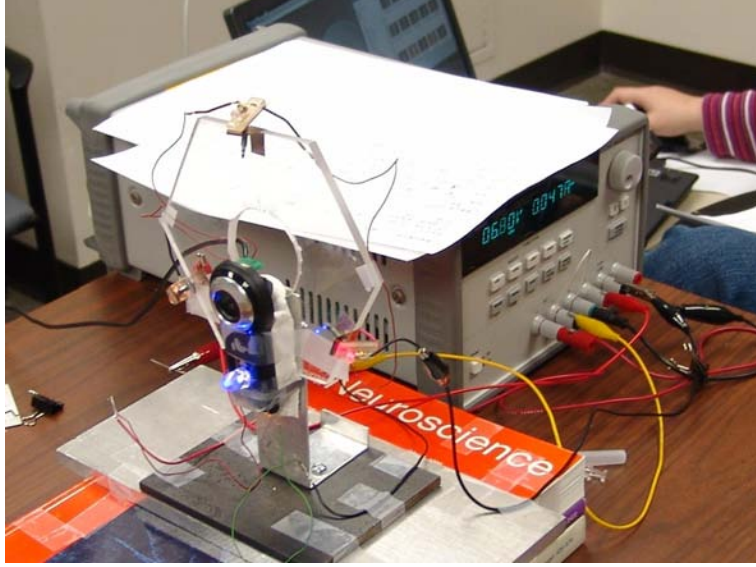


Figure 38: Prototype with a miniature camera.

We re-calibrated the laser beams and the camera with the method discussed in Chapter 5. The spatial equations of all three laser beams were estimated as

$$\begin{array}{rcl} \frac{x+42.6925}{0.2018} & = & \frac{y+71.9837}{0.1024} = \frac{z-329.4578}{0.9741} \\ \frac{x-13.2503}{-0.2334} & = & \frac{y-10.2144}{-0.1274} = \frac{z-329.6447}{0.9640} \\ \frac{x+28.0141}{-0.0170} & = & \frac{y+60.9772}{0.3809} = \frac{z-329.4865}{0.9245} \end{array} \quad (7.3)$$

The camera intrinsic matrix \mathbf{K} in the pixel units was found to be

$$\mathbf{K} = \begin{bmatrix} 1313.26058 & 0 & 774.89935 \\ 0 & 1328.44812 & 600.44398 \\ 0 & 0 & 1 \end{bmatrix}$$

The camera pitch size was $2.8 \mu m$ as provided in the specification sheets and the measured fixed focal length was 3.7 mm. The measured rotation matrix and translation vectors were

$$\mathbf{R} = \begin{bmatrix} 0.997709 & -0.004644 & -0.067489 \\ -0.011712 & -0.994433 & -0.104716 \\ -0.066627 & 0.105266 & -0.992210 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 50.063508 \\ 214.271946 \\ 528.517606 \end{bmatrix}$$

Our algorithm was tested by placing a paper board parallel with the camera. Seven real objects were placed at 20, 50, 90 degrees to the horizontal plane with different distances (shown in Figure 39). The estimation results of the dimension and height are listed in Table 3 to Table 7.

Table 3: Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 20° with respect.

	Cookie	Hamburger	Ball	pizza
Number of Trials	23	17	15	9
Mean and STD of Absolute Error of Length (mm)	2.92 ± 3.41	2.52 ± 2.30	7.75 ± 3.38	14.2 ± 18.3
Mean and STD of Relative Error of Length (%)	4.49 ± 5.26	4.20 ± 3.84	19.30 ± 6.64	14.2 ± 18.3

Table 4: Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 20° with respect (cont.).

	Box	Bread	Cup
Number of Trials	30	25	10
Mean and STD of Absolute Error of Length (mm)	1.82 ± 1.91	3.80 ± 4.22	2.78 ± 3.15
Mean and STD of Absolute Error of Height (mm)	1.02 ± 0.74	1.71 ± 1.76	4.85 ± 6.07
Mean and STD of Relative Error of Length (%)	2.69 ± 2.81	3.92 ± 4.36	4.49 ± 5.09
Mean and STD of Relative Error of Height (%)	4.88 ± 3.51	13.2 ± 13.5	4.05 ± 5.07

Table 5: Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 50° with respect.

	Cookie	Hamburger	Ball	pizza
Number of Trials	23	17	15	9
Mean and STD of Absolute Error of Length (mm)	2.38 ± 2.76	1.42 ± 1.78	7.68 ± 3.64	9.32 ± 4.50
Mean and STD of Relative Error of Length (%)	3.66 ± 4.26	3.37 ± 2.96	19.2 ± 9.1	9.32 ± 4.50

Table 6: Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 50° with respect (cont.).

	Box	Bread	Cup
Number of Trials	30	25	10
Mean and STD of Absolute Error of Length (mm)	2.44 ± 2.36	3.50 ± 3.14	1.94 ± 3.05
Mean and STD of Absolute Error of Height (mm)	6.30 ± 2.65	1.92 ± 1.28	23.9 ± 10.00
Mean and STD of Relative Error of Length (%)	3.59 ± 3.46	3.61 ± 3.25	3.14 ± 4.93
Mean and STD of Relative Error of Height (%)	30 ± 12.6	14.82 ± 9.88	19.9 ± 8.38

Table 7: Dimension and height estimates using laser structured lights at different distances. The degree of rotation of the object was 90° with respect.

	Paper Board	Cookie	Box	Bread
Number of Trials	26	23	30	25
Mean and STD of Absolute Error of Length (mm)	2.35 ± 1.35	1.88 ± 1.83	6.03 ± 3.91	4.10 ± 2.47
Mean and STD of Relative Error of Length (%)	5.87 ± 3.39	2.90 ± 2.80	8.87 ± 5.75	4.24 ± 2.55

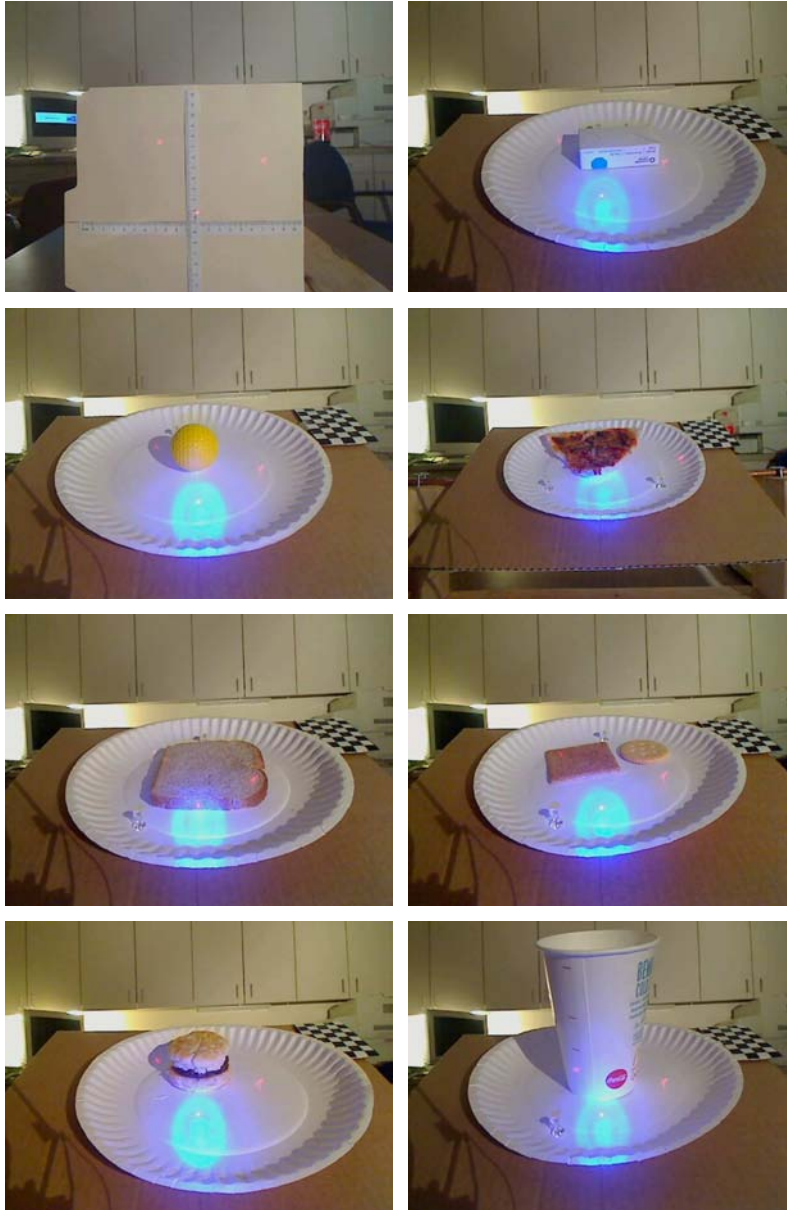


Figure 39: Real objects used for dimension and height estimation

7.2 LED BEAM DESIGN EXPERIMENT

Three experiments were performed to evaluate the feasibility and performance of the LED spotlight design. The device prototype is shown in Figure 38. The first experiment was to estimate the distance while keeping the object plane perpendicular to the camera optical axis. The second experiment was to estimate both the oblique angle and distance for a tilted object plane. The third experiment was to estimate the dimension and height of eight objects (shown in Figure 39). The intrinsic parameters of the camera were calibrated before the experiments. In each image, the center position and dissimilarity of the spotlight pattern were extracted after segmentation and ellipse fitting. The experimental results are listed in Tables 8 through 11.

Table 8: Estimated distances from perpendicular object plane

Distance l (mm)	20	25	30	35	40	45
Estimated l (mm)	20.65	25.21	29.97	34.71	40.14	45.22

Table 9: Estimated distances and angles from tilted object planes

True Estimated	$l=30\text{cm}$			$l=40\text{cm}$		
	$\theta=30^\circ$	$\theta=40^\circ$	$\theta=50^\circ$	$\theta=30^\circ$	$\theta=40^\circ$	$\theta=50^\circ$
l (cm)	29.10	28.87	28.95	38.33	38.46	38.33
θ ($^\circ$)	24.3	36.2	48.3	23.4	41.7	51.9

Table 10: Dimension and height estimates using LED structured light at different distances and a plane tilted by 20°

	Box	Bread	Cookie	Hamburger
Number of Trials	10	10	10	10
Mean and STD of Absolute Error of Length (mm)	1.95 ± 2.29	13.15 ± 25.48	7.95 ± 6.10	0.88 ± 0.70
Mean and STD of Relative Error of Length (%)	2.87 ± 3.37	13.56 ± 26.27	12.62 ± 9.69	2.95 ± 2.35
Mean and STD of Absolute Error of Height (mm)	1.96 ± 1.16	4.72 ± 1.58	0.55 ± 0.71	4.62 ± 2.55
Mean and STD of Relative Error of Height (%)	9.33 ± 5.51	47.24 ± 15.80	13.75 ± 17.87	13.06 ± 7.5

Table 11: Dimension and height estimates using LED structured light at different distances and a plane tilted by 50°

	Box	Bread	Cookie	Hamburger
Number of Trials	10	10	10	10
Mean and STD of Absolute Error of Length (mm)	-2.87 ± 2.62	11.91 ± 23.39	-7.82 ± 4.55	-1.50 ± 0.72
Mean and STD of Relative Error of Length (%)	-4.22 ± 3.86	12.28 ± 24.12	-12.4 ± 7.23	-4.99 ± 2.43
Mean and STD of Absolute Error of Height (mm)	2.39 ± 2.33	1.17 ± 1.24	0.43 ± 1.57	-0.56 ± 5.79
Mean and STD of Relative Error of Height (%)	11.40 ± 11.68	12.37 ± 15.80	10.65 ± 39.27	-1.66 ± 17.04

7.3 DISCUSSION

In this chapter, we have tested the proposed prototypes and dimension estimation algorithms by performing experiments with various artificial objects and real foods. Cameras with both high and median resolutions have been used as image acquisition instruments of the system.

For the laser beam design, the distance and physical dimension estimates with a higher resolution camera show better performance because the higher resolution reduced image processing errors in system calibration and feature extraction. The relative average error with a high resolution camera is less than 10 % (shown in Table 1 and Table 2). With a median resolution camera, both the dimensions in the object plane and the height of the object are estimated at configurations with three angles (20° , 50° , and 90° to the horizontal plane) and multiple distances. For objects with a cubic shape, the dimension estimates have an average relative error of less than 10%. The average errors are similar for other test objects at all angles. The results show the robustness of our estimation algorithms. The edges of objects with certain shapes (e.g., a slice of pizza) are more difficult to extract, which lead to bigger estimation errors. The height estimation was tested with three objects: a box, a slice of bread, and a paper cup. The absolute average errors for the box and bread slice are approximately 2mm and 3.5mm at 20° tilted plane, 9mm and 3mm at 50° tilted plane, and 10mm and 6mm at 90° tilted plane, respectively. Our comparative study shows that the height estimation algorithm performs better at smaller angles, and the dimension estimation algorithms achieve better accuracy at angles between 20° and 50° . Because the shape of the cup is not considered a regular shape, the estimation error is relatively larger.

For the LED spotlight design, our algorithms estimated the distance from the camera to a parallel object plane, the tilted angle of object plane with respect to the horizontal plane, and the object dimensions. The distance estimation from the perpendicular object plane has an absolute average error of less than 1mm for objects at different positions (Table 8). This error is mainly caused by the system calibration error and the image processing error. The distance estimation in tilted object planes shows different performances (Table 9). The larger the tilted angle, the larger the estimation error. This is likely caused by uneven attenuation on the edges of the LED spotlight resulting in bias at the elliptic center. For a

larger tilted angle, the differences of intensity attenuation in different directions are larger. The dimensional and height estimates of five objects (Table 10 and 11) in planes with tilted angles between 20 and 50 degrees show similar absolute and relative errors.

Our findings are summarized as follows:

- The laser beam method works well on cubic objects at a wide range of locations. However, the energy requirement, space, and cost of this design is higher than the LED design;
- The LED spotlight method works better when the object plane is parallel (or almost parallel) to the camera plane than at largely tilted positions;
- The standard deviations are both relatively large because the number of trials is not sufficiently large;
- For cubic objects, both methods work well when the rotation angle of the tilted plane is between 20° and 50° ;
- For non-cubic objects, height estimates have larger errors because of the difficulty in finding the points that define height.

8.0 CONCLUSION

An accurate food dimension estimation method with a portable and affordable device is highly desirable in dietary assessment for obesity study and treatment. In this dissertation, we have presented two newly designed prototypes, a laser beam structured lights prototype and an LED spotlight prototype. Novel methods based on geometric perspective projection have been presented to accurately estimate 3D physical dimensions of food from a single image. The proposed methods successfully estimated food dimensions in arbitrary positions with the help of simple structured lights. The flowchart of the proposed food dimension estimation approach is described in Figure 40.

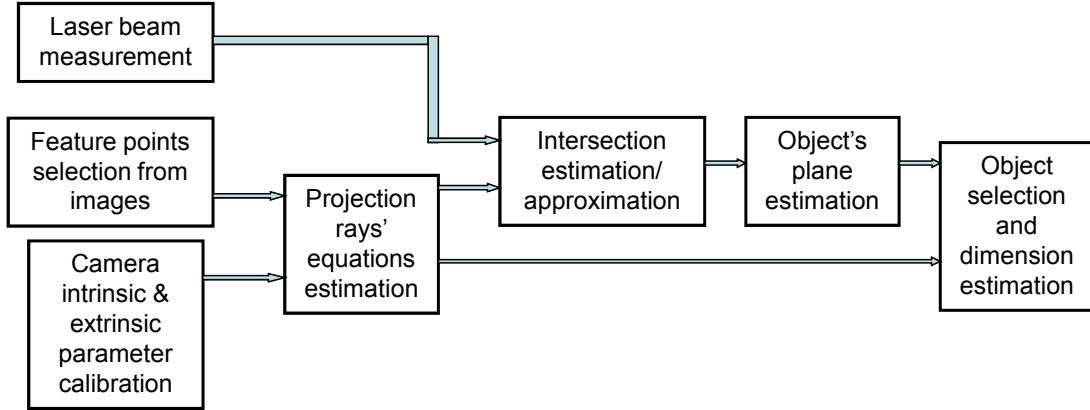


Figure 40: Flowchart of the proposed food dimension estimation approach

Two geometric algorithms have been developed, which were based on perspective projection geometry and optical triangulation. The first algorithm utilized laser beam structured lights and was based on an ideal model of equilateral triangular structured lights. The plane containing three reflected laser spots was determined by the geometry of the pinhole camera

and all three laser beams. To handle the difficulty in setting up a very precise equilateral triangle with three laser beams, an approach to individually calibrate and reconstruct the laser beams was developed. An orthogonal linear regression method was used to reconstruct the paths of the laser beams. Due to errors from system calibration, feature extraction, and image digitalization, the approximated object plane was estimated using a modified algorithm. The experimental results verified the performance of the proposed methods in estimating dimensions with both artificial and real foods, in which a relative average error less than 10% was achieved.

The second algorithm utilized an LED structured light based on the circular feature's deformation caused by the rotation, transformation, and reflection of a planar surface. The image of LED spot was described by a feature vector, including the ratio of major and minor axes, the position of the center, and the skew of the ellipse. The deformation of structured light pattern was modeled as a function of the location of the object plane. The spatial position of the plane was determined using the feature vector obtained in the feature extraction phase, and the dimensions of the objects were then calculated from it. This algorithm currently worked well with object planes rotated in a single direction and it was more sensitive to the error in rotation angle estimation than the laser beams design.

A height estimation algorithm was also developed for regularly shaped objects in order to estimate food volume. Accuracy of this algorithm depended on two factors: the accuracy in the object plane estimation and the accuracy in height segmentation. The limitation of this algorithm is that it depends on the visibility of the object's height in the image.

To extract the structured spotlight from images, we applied a global thresholding method to segment the spotlight and used an ellipse fitting method. After investigating the color properties of the laser beam and the LED spotlight, we selected the difference-image and the green-channel as the inputs for segmentation for the two structured light designs, respectively. A histogram-based thresholding method and a clustering-based thresholding method were applied. A comparative study with these two methods was conducted, and the former method showed better performance on our image data sets. The Sobel edge detector was used to detect the edges of spotlight after segmentation. The detected edges were discontinuous and noisy. One reason was that the laser spots and the LED spots possessed blurred edges

in the acquired images because of the natural divergence of the light beams. Moreover, the reflective surface was usually not smooth, so the intensities of the regions close to the boundaries were uneven. Since the projected patterns were known to be circular, we applied an ellipse fitting algorithm to compensate for the noise effect.

Prototypes have been developed and are ready for the clinical tests [\[117\]](#). We believe the proposed approaches and techniques have many other potential applications in medical and non-medical fields.

9.0 FUTURE WORK

9.1 MORE EFFICIENT SYSTEM DESIGN AND FEATURE EXTRACTION DIFFICULTIES

In this dissertation, several algorithms based on structured laser beams prototype have shown satisfactory performance to determine planar patch dimensions accurately and efficiently. However, the laser beams prototype requires more power and space to place and drive three laser modules. Under some conditions, we cannot detect all three highlights from the captured images because of occlusion, reflection, object's color and other uncertainties. One laser beam with a diffraction grating film is a more efficient in design and can provide multiple reference points (such as a point array). However, the feature extraction of uneven illumination intensity is still a very challenging image processing problem. A global thresholding segmentation technique, such as that used in this dissertation, will no longer work in the given situation. An adaptive local thresholding method is more suitable for the design with attenuating point array pattern. Therefore, we will pursue the development of good adaptive segmentation methods with respect to the characteristics of particular features in our future work.

The proposed structured lights design and their corresponding estimation algorithms assume that the laser and LED highlighted regions are elliptic and isotropically attenuating. However, the laser highlights and the LED light spots present different intensities on the closer edges and further edges in real images. This is caused by the different travel distances of the light. This property results in difficulties in accurately extracting the regions of highlighted areas. Currently, we use same thresholds to segment spots in every directions. Since good segmentation and shape analysis of the structured lights in images directly determine

geometric parameters of the algorithm, we will develop new algorithms to improve performance of the segmentation for varying attenuation in different direction of the laser and LED highlights.

9.2 ROBUST DIMENSION ESTIMATION

We currently use a traditional *optical triangular method* to estimate the 3D coordinates of a point based on the perspective model of a pinhole camera. The algebraic solution of this model is neat and easy to be implemented. However, this method suffers from system calibration errors and image processing noise. In our future work, we will develop robust methods with better tolerance to such noises and uncertainties to improve the performance of the 3D dimension estimation.

10.0 PUBLICATIONS

- [1] M. Sun, J. Fernstrom, W. Jia, S. Hackworth, **N.Yao**, Y. Li, C. Li, M. Fernstrom, and R. J. Sclabassi, *A wearable electronic system for objective dietary assessment*, Journal of the American Dietetic Association, Vol. 110(1), pp. 45-47, Jan. 2010.
- [2] **NingYao**, Qiang Liu, Robert J. Sclabassi, and Mingui Sun, *Sparse Representation of Physical Activity Video in the Study of Obesity*, IEEE International Symposium on Circuits and Systems, pp. 2582-2585, May 18-21, 2008.
- [3] **NingYao**, Ruizhen Zhao, Robert J. Sclabassi, Mingui Sun *Aided Structured Light in Planar Object Physical Dimension Measurement*, IEEE 34rd Annual Northeast Bioengineering Conference, pp. 298-299, March 2008.
- [4] **NingYao**, R. J. Sclabassi, Qiang Liu, Jie Yang, John D. Fernstrom, Madelyn H. Fernstrom, Mingui Sun, *A Video Processing Approach to the Study of Obesity*, 2007 IEEE International Conference on Multimedia and Expo, pp. 1727-1730, July 2007
- [5] **NingYao**, Sclabassi, R. J. Qiang Liu, Mingui Sun, *A video-based algorithm for food intake estimation in the study of obesity*, IEEE 33rd Annual Northeast Bioengineering Conference, pp. 298-299, March 2007.
- [6] **NingYao**, Heung-No Lee, Cheng-Chun Chang, Sclabassi, R.J. Sclabassi, Mingui Sun, *A Power-Efficient Communication System between Brain-Implantable Devices and External Computers*, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6588-6591, Aug. 2007.
- [7] **NingYao**, Heung-No Lee, R.J. Sclabassi, Mingui Sun, *Low Power Digital Communication in Implantable Devices Using Volume Conduction of Biological Tissues*, 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6249-6252, 2006.

- [8] **NingYao**, Man-Wai Kwan, Chi-Wah Kok, *Correlation-based adaptive filters for channel identification*, IEEE International Symposium on Circuits and Systems, Vol. 4, pp. 3737-3740, May 2005.
 - [9] **NingYao**, Man-Wai Kwan Fung, C.C. Chi-Wah Kok, *Higher-Order Statistics Based Iterative Space-Time FIR Precoder-Blind Equalizer*", The Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, pp. 1024-1028, 2005.
- Other publications:**
- [9] Mingui Sun,; Qiang Liu, Karel Schmidt, Jie Yang, **NingYao**, John D. Fernstrom, Madelyn H. Fernstrom, James P. DeLany, Robert J. Sclabassi, *Determination of food portion size by image processing*, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 871-874, 20-25 Aug. 2008.
 - [10] Hong Zhang,Kui Zhang, **NingYao** ,Robert J. Sclabassi, Mingui Sun, *Refined segmentation of images for human pose analysis* *Engineering in Medicine and Biology Society*, the 30th Annual International Conference of the IEEE, pp. 4809-4811, 20-25 Aug. 2008.
 - [11] Kui Zhang, Hong Zhang, **NingYao**, Robert J. Sclabassi, Mingui Sun, *Carried Load Measurement Based on Gait Analysis and Human Kinetics*, Congress on Image and Signal Processing(CISP), Vol. 3, pp. 104-107, 27-30 May 2008.
 - [12] Lu Li, Hong Zhang, **NingYao**, *Enforcing control points of correspondence by the Hausdorff distance*, International Conference on Information and Automation(ICIA), pp. 1778-1782, June 2008.
 - [13] Qiang Liu, Sclabassi, R.J., **NingYao**, Mingui Sun, *3D Construction of Endoscopic Images Based on Computational Stereo* , IEEE 33rd Annual Northeast Bioengineering Conference, pp. 69 - 70, April 2006.

APPENDIX

GRADIENT DESCENT SEARCHING METHOD

We propose a gradient descent searching method to find two points on each of two skew lines, which have the smallest distance between two lines. The middle-point of these two points is used as the approximated intersection in our method.

Table 12: Searching the approximated intersection point between two skew lines

Definition:

l and l' : Two skew lines,

p , and p' : Point on line l and line l' ,

\hat{d} = $d(\hat{p}, \hat{p}')$, shortest distance between two skew lines l and l' , known,

$d(p, l')$ = $\min d(p, p'), p' \in l'$, shortest distance between a point p and a line l' ,

μ : Step-size with $0 < \mu < 1$,

ϵ : Stop criterion.

There are two steps to approximate the intersection point of two skew lines.

Step One: searching the closest point \hat{p} to line l' .

Iteration (for $n = 1, 2, \dots, N$), while $e(n) < \epsilon$,

$$d(n) = d(p_n, l'),$$

$$e(n) = d(n) - \hat{d},$$

p_{n+1} = translate p_n on l alone positive z direction by $e(n) \cdot \mu$

$$d(n+1) = d(p_{n+1}, l'),$$

If $d(n+1) > d(n)$, p_{n+1} = translate p_n on l alone negative z direction by $e(n) \cdot \mu$

Table 13: Searching the approximated intersection point between two skew lines (Continued)

Step Two: searching the closest point \hat{p}' to point \hat{p} .

Iteration (for $m = 1, 2, \dots, M$), while $e(m) < \epsilon$,

$$d(m) = d(p'_m, \hat{p}),$$

$$e(m) = d(m) - \hat{d},$$

p'_{m+1} = translate p'_m on l' alone positive z direction by $e(m) \cdot \mu$

$$d(m+1) = d(p'_{m+1}, \hat{p}),$$

If $d(m+1) > d(m)$, p_{m+1} = translate p_m on l alone negative z direction by $e(m) \cdot \mu$

BIBLIOGRAPHY

- [1] <http://www.healthymamericans.org>, Trust for American Health, Robert Wood Johnson Foundation.
- [2] Henrik Møller, Anders Møllegaard, Knud Lindvig and Jørgen H. Olsen, *Obesity and cancer risk: a danish record-linkage study*, European Journal of Cancer, Vol. 30(3), pp. 344-350, 1994.
- [3] Frank B. Hu, JoAnn E. Manson, Meir J. Stampfer, Graham Colditz, Simin Liu, Caren G. Solomon and Walter C. Willett, *Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women*, The New England Journal of Medicine, Vol. 345, pp. 790-797, Sep., 2001.
- [4] H. B. Hubert, M. Feinleib, P. M. McNamara and W. P. Castelli, *Obesity as an independent risk factor for cardiovascular disease: a 26- year follow-up of participants in the Framingham Heart Study*, Circulation by American Heart Association, Vol. 67, pp. 968-977, 1983.
- [5] Roland T. Jung, *Obesity as a disease*, Oxford Journals, Medicine, British Medical Bulletin, Vol. 53(2), pp. 307-321, 1997.
- [6] Seung-Han Suk, Ralph L. Sacco, Bernadette Boden-Albala, Jian F. Cheun, John G. Pittman, Mitchell S. Elkind and, Myunghee C. Paik, *Abdominal Obesity and Risk of Ischemic Stroke*, The Northern Manhattan Stroke Study, Stroke, Vol. 34, pp. 1586-1592, 2003.
- [7] C. S. Ray, D. Y. Sue, G. Bray, J. E. Hansen and K. Wasserman, *Effects of obesity on respiratory function*, Am Rev. Respir. Dis., Vol. 128, pp. 501-506, 1983.
- [8] D. T. Felson, J. J. Anderson, A. Naimark, A. M. Walker and R. F. Meenan, *Obesity and knee osteoarthritis. The Framingham Study.*, Ann. Intern. Med., Vol. 109(1), pp. 18-24, Jul. 1988.
- [9] http://www.surgeongeneral.gov/topics/obesity/calltoaction/fact_consequences.htm
- [10] Margaret L. Watkins, Sonja A. Rasmussen, Margaret A. Honein, Lorenzo D. Botto and Cynthia A. Moore, *Maternal Obesity and Risk for Birth Defects*, Pediatrics, Vol. 111(5), pp. 1152-1158, May 2003.

- [11] James O. Hill and John C. Peters, *Environmental Contributions to the Obesity Epidemic*, Science, Vol. 280(5368), pp. 1371-1374, May 1998.
- [12] E. V. Villanueva, *The validity of self-reported weight in US adults: a population based cross-sectional study*, BMC Public Health, pp. 1-11, 2001.
- [13] <http://www.calorieking.com/>
- [14] <http://www.mobigloo.com/software/caloriecounter/>
- [15] JR. W. Jeffery, *Bias in reported body weight as a function of education, occupation, health and weight concern*, Addict. Behav., Vol. 21, pp. 217 -222, 1996.
- [16] S. Stotland and M. Larocque, *Convergent validity of the Larocque Obesity Questionnaire and selfreported behavior during obesity treatment*, Psychological Report, Vol. 95, pp. 1031-1042, 2004.
- [17] R. J. Hill and P. S. W. Davies *The validity of self-reported energy intake as determined using the doubly labelled water technique*, British Journal of Nutrition, Cambridge University Press, Vol. 85(4), pp. 415-430, Apr. 2001.
- [18] C. J. Boushey, *Improving diet and physical activity assessment*, RO1, 2005.
- [19] I. Woo, K. Otsmo, S. Kim, D. Ebert, E. Delp and C. Boushey, *Automatic portion estimation and visual refinement in mobile dietary assessment*, Proc. SPIE, Computational Imaging VIII, Vol. 7533, 75330O, Jan. 2010.
- [20] A. B. Koc, *Determination of watermelon volume using ellipsoid approximation and image processing*, Postharvest Biology and Technology, Vol. 45, pp. 366-371, Sep. 2007.
- [21] M. Khojastehnazhand, M. Omid and A. Tabatabaeefar, *Determination of orange volume and surface area using image processing technique*, International Agrophysics, Vol. 23(3), pp. 237-242, 2009.
- [22] M. Rashidi and M. Gholami, *Determination of kiwifruit volume using ellipsoid approximation and image-processing methods*, Int. J. Agri. Biol., Vol. 10, pp. 375-380, 2008.
- [23] H. Zhang, K. Y. K. Wong and G. Zhang, *Camera calibration from images of spheres*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29(3), pp. 499-502, 2007.
- [24] X. Ying and H. Zha, *A novel linear approach to camera calibration from sphere images*, International Conference on Pattern Recognition, Vol. 1, pp. 535-538, 2006.
- [25] C. Yu and Q. Peng, *Robust recognition of checkerboard pattern for camera calibration*, Optical Engineering, Vol. 45(9), 2006.

- [26] D. Nister, O. Naroditsky and J. Bergen, *Visual odometry*, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 652-659, 2004.
- [27] M. Puri, Zhiwei Zhu, Qian Yu, A. Divakaran, H. Sawhney, *Recognition and volume estimation of food intake using a mobile device*, Workshop on Applications of Computer Vision (WACV), pp. 1-8, Dec. 2009.
- [28] G. Pingali and R. Jain, *Electronic Chronicles: Empowering Individuals, Groups, and Organisations*, IEEE International Conference on Multimedia and Expo, pp. 1540-1544, 2005.
- [29] R. Jain, *Media Vision: A True Multimedia Client*, IEEE Multimedia, April 2005.
- [30] Jon Radoff, <http://radoff.com/blog/2008/08/22/anatomy-of-an-mmorpg/>
- [31] Betsy Schiffman, http://www.wired.com/techbiz/media/news/2008/04/3d_movies
- [32] Eden Ashley Umble *Making it real: the future of stereoscopic 3D film technology*, ACM SIGGRAPH, Computer Graphics, Vol. 40(1), May 2006.
- [33] Patric Maynard, *Drawing distinctions: the varieties of graphic expression*, Cornell University Press, pp. 22, 2005.
- [34] B. K. P. Horn, *Height and gradient from shading*, International J. of Computer Vision, Vol. 5,(1), pp. 37-75, 1990.
- [35] B. K. P. Horn. *Shape from shading : a method for obtaining the shape of a smooth opaque object from one view*, Ai-tr-232, MIT, 1970.
- [36] Y. Uranishi, M. Naganawa, Y. Yasumuro, M. Imura, Y. Manabe, K. Chihara, *Whole Shape Measurement System Using a Single Camera and a Cylindrical Mirror*, Pattern Recognition, 18th International Conference on Vol. 4, pp. 866 - 869, 2006.
- [37] R. White, D. A. Forsyth *Combining Cues: Shape from Shading and Texture* IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 1809-1816, 2006.
- [38] R. Zhang, P. Tsai, J. Cryer, and M. Shah, *Shape from shading: A survey* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21(8), pp. 690-706, 1999.
- [39] P. Parodi, G. Piccioli, *3D shape reconstruction by using vanishing points* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18(2), pp. 211-217, 1996.
- [40] C.K. Fong, W.K. Cham, *3D object reconstruction from single distorted line drawing image using vanishing points*, Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems, pp. 53-56, Dec. 2005.

- [41] Yong-In Yoon, Jang-Hwan Im, Jong-Soo Choi, Jin-Tae Kim, Dong-Wook Kim, Jun-Sik Kwon, *Reconstruction of linearly parameterized models three vanishing points from a single image of perspective projection*, Proceedings. Seventh International Symposium on Signal Processing and Its Applications, Vol. 1, pp. 13-16, July 2003.
- [42] D. Jelinek, C.J. Taylor, *Reconstruction of linearly parameterized models from single images with a camera of unknown focal length* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, Issue 7, pp. 767-773, July 2001.
- [43] D. Hoiem, A. Efros, and M. Herbert, *Geometric context from a single image*, in International Conference on Computer Vision, Vol. 1, pp. 654-661, 2005.
- [44] M. Subbarao, *Direct recovery of depth-map I: Differential methods*, in Proc. IEEE Comput. Soc. Workshop Comput. Vision, 1987, pp. 58-65.
- [45] A. Heyden and M. Pollefeys, *Multiple View Geometry* Emerging Topics in Computer Vision, G. Medioni and S.B. Kang, eds., Prentice Hall, chapter 3, pp. 45-108, 2003.
- [46] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049, 2000.
- [47] <http://www.wikipedia.org>
- [48] O. Faugeras, *Three-dimensional Computer Vision : A Geometric Viewpoint*, MIT Press, 1993.
- [49] R. Y. Tsai, *Multiframe image point matching and 3D surface reconstruction*, IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP), pp. 1272-1275, 1983.
- [50] R. Y. Tsai, *Synopsis of recent progress on camera calibration for 3D machine vision*, The Robotics Review, pp. 147-159, 1989.
- [51] R. Y. Tsai, *A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses* IEEE J. Robotics and Automation, Vol. 3(4), pp. 323-344, Aug. 1987.
- [52] W. Faig, *Calibration of close-range photogrammetry systems: Mathematical formulation*, Photogrammetric Eng. Remote Sensing, Vol. 41, pp. 1479-1486, 1975.
- [53] Y. I. Abdel-Aziz and H. M. Karara *Direct linear transformation into object space coordinates in close-range photogrammetry*, in Proc. Symp. Close-Range Photogrammetry, Univ. of Illinois at Urbana-Champaign, Urbana, IL, pp. 1-18, 1977.
- [54] Y. I. Abdel-Aziz and H. M. Karara *Photogrammetric potential of non-metric camera*, Civil Engineering Studies, Photogrammetry Series 36. Univ. of Illinois, Urbana, IL, 1974.

- [55] Z. Zhang, *A flexible new technique for camera calibration*, Microsoft Technical Report, MSR-TR-98-71.
- [56] P. Sturm and S. Maybank, *On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 432-437, June 1999.
- [57] S. J. Maybank and O. D. Faugeras, *A Theory of Self-Calibration of a Moving Camera* Intl J. Computer Vision, Vol. 8(2), pp. 123-152, Aug. 1992.
- [58] <http://www.cs.jhu.edu/~wolff/course600.461/week9.3/index.htm>
- [59] A. Criminisi, I. Reid and A. Zisserman, *Single view metrology*, International Journal of Computer Vision, Vol. 40(2), pp. 123-148, 2000.
- [60] Brian Curless, *From Range Scans to 3D Models*, ACM SIGGRAPH Computer Graphics, Vol. 33(4), pp. 38-41, Nov 2000.
- [61] Y. Sun, J.K. Paik, A. Koschan, M.A. Abidi, *3D reconstruction of indoor and outdoor scenes using a mobile range scanner*, in Procs. 16th International Conference on Pattern Recognition, Vol. 3, pp. 653-656, 2002.
- [62] Song Zhang, Peisen Huang, *High-resolution, real-time 3-D shape measurement*, Optical Engineering, Vol. 45(12), pp. 123601(1-8), Dec. 2006.
- [63] J. Heikkila, O. Silven, *A four-step camera calibration procedure with implicit image correction*, Proc. on Computer Vision and Pattern Recognition, pp. 1106-1112, 1997.
- [64] J.G. Fryer and D.C. Brown, *Lens Distortion for Close-Range Photogrammetry*, Photogrammetric Engineering and Remote Sensing Vol. 52(1) pp. 51-58, 1986.
- [65] Gang Xu, Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified approach*, Kluwer Academic Publishers, 1996.
- [66] M. Pilu, *A Direct Method for Stereo Correspondence Based on Singular Value Decomposition*, Proc. IEEE Computer Vision and Pattern Recognition Conf., pp. 261-266, 1997.
- [67] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, *A comparison and evaluation of multi-view stereo reconstruction algorithms*. CVPR, Vol. 1, pp. 519-528, 2006.
- [68] A. Fusiello, E. Trucco, and A. Verri, *A compact algorithm for rectification of stereo pairs*, Machine Vision and Applications, Vol. 12(1), pp. 16-22, 2000.
- [69] R. Hartley, *Theory and practice of projective rectification*, International Journal of Computer Vision, Vol. 35(2), pp. 1-16, 1999.

- [70] C. Loop, and Z. Zhang, *Computing rectifying homographies for stereo vision*, CVPR, Fort Collins, CO, pp. 125-131, 1999.
- [71] M. Pollefeys, R. Koch, and L. VanGool, *A simple and efficient rectification method for general motion*, International Conference on Computer Vision, Corfu, Greece, pp. 496-501, 1999.
- [72] D. Marr and T. Poggio, *Cooperative computation of stereo disparity*, Science, Vol. 194(4262), pp. 283-287, 1976.
- [73] F. Blais *Review of 20 years of range sensor development*, Journal of Electronic Imaging, Vol. 13(1), pp. 231-240, 2004.
- [74] G. Chazan and N. Kiryati, *Pyramidal Intensity-Ratio Depth Sensor*, Tech. Rep. No. 121, Israel Institute of Technology, Technion, Haifa, Israel, 1995.
- [75] F.M. Henderson, and A.J. Lewis, Principles and applications of imaging radar. Manual of remote sensing: Third edition, Vol. 2, John Wiley and Sons, Inc., 1998.
- [76] Daniel Scharstein and Richard Szeliski, *High-Accuracy Stereo Depth Maps Using Structured Light*, In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), Vol. 1, pp. 195-202, Madison, WI, June 2003.
- [77] Ashutosh Saxena, Min Sun, Andrew Y. Ng., *Learning 3-D Scene Structure from a Single Still Image*, In ICCV workshop on 3D Representation for Recognition (3dRR-07), 2007.
- [78] Ashutosh Saxena, Min Sun, Andrew Y. Ng., *3-D Reconstruction from Sparse Views using Monocular Vision*, In ICCV workshop on Virtual Representations and Modeling of Large-scale environments (VRML), 2007.
- [79] Ashutosh Saxena, Min Sun, Andrew Y. Ng., *Make3D: Learning 3-D Scene Structure from a Single Still Image*, in IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31(5), pp. 824-840, May 2009.
- [80] K. Grauman, G. Shakhnarovich, and T. Darrell, *Inferring 3D structure with a statistical image-based shape model*, Ninth IEEE International conference on Computer Vision, Vol. 1, pp. 641-647, 2003.
- [81] V. Aslantas, *A depth estimation algorithm with a single image*, Optics Express, Vol. 15(8), pp. 5024-5029, 2007.
- [82] Shang-Hong Lai, Chang-Wu Fu, Shyang Chang, *A Generalized Depth Estimation Algorithm with a Single Image*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14(4), pp. 405-411, 1992.
- [83] M. Subbarao, and T.S. Choi, *Accurate recovery of three dimensional shape from image focus*, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 266-274, March 1995.

- [84] M. Subbarao, T.S. Choi, and A. Nikzad, *Focusing Techniques*, Journal of Optical Engineering, pp. 2824-2836, Nov. 1993.
- [85] S.K. Nayar and Y. Nakagawa, *Shape from Focus*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16(8), pp. 824-831, Aug. 1994.
- [86] M. Subbarao and J.K. Tyan, *Selecting the optimal focus measure for auto-focusing and depth-from-focus*, IEEE Trans Pattern Anal Machine Intell Vol. 20, pp. 864-870, 1998.
- [87] Yen-Fu Liu, *A unified approach to image focus and defocus analysis*, Ph.D Dissertation. State University of New York at Stony Brook, 1998.
- [88] Paolo Favaro and Stefano Soatto, *Learning Shape from Defocus*, European Conference on Computer Vision, pp. 735-745, 2002.
- [89] Yoav Y. Schechner and Nahum Kiryati, *Depth from defocus vs. stereo: How different really are they?*, International Journal of Computer Vision, Vol. 89, pp. 141-162, 2000.
- [90] Thomas Lu and Tien-Hsin Chao, *A single-camera system captures high-resolution 3D images in one shot*, Electronic Imaging and Signal Processing, SPIE Newsroom, 28 November 2006.
- [91] Anat Levin Rob Fergus Fredo Durand William T. Freeman, *Image and depth from a conventional camera with a coded aperture*, International Conference on Computer Graphics and Interactive Techniques, Article No. 70, 2007.
- [92] Jean-Yves Bouguet, Camera Calibration Toolbox for Matlab.
- [93] A P. Pentland, *A new sense for depth of field*, IEEE Trans. Patt. Anal. Machine Intell., Vol. 9, pp. 523-531, 1987.
- [94] P. Grossmann, *Depth from focus*, Patt. Recogn. Lett., Vol. 5(1), pp. 63-69, 1987.
- [95] E.S. McVey, and G.L. Jarvis, *Ranking of patterns for use in automation*, IEEE Trans. Ind. Electron Contr. Instrum., Vol. IECI-24(2), pp.211-213, May, 1977.
- [96] R.A.Young, *Locating industrial parts with subpixel accuracies*, SPIE Proc. Optics, Illumination, Image Sensing Machine Vision, Vol. 728, pp. 2-9, Oct. 1986.
- [97] R. Safaee-Rad, I. Tchoukanov, K.C. Smith, B.Benhbab, *Three-dimensional location estimation of circular features for machine vision*, IEEE Trans. on Robotics and Automation, Vol. 8(5), pp. 624-640, Oct. 1992.
- [98] M.J. Magee, and J.K. Aggarwal, *determining the position of a robot using a single calibratino object*, in Proc. IEEE Int. Conf. Robotics Automat., pp. 140-149, Mar, 1984.
- [99] M.R. Kabuka and A.E. Arenas, *Position verification of a mobile robot using standard pattern*, IEEE J. Robotics Automat., Vol. RA-3(6), pp. 505-516, Dec. 1987.

- [100] B. Hussain and M.R. Kabuka, *Real-time system for accurate three-dimensional position determination and verification*, IEEE Trans. Robotics Automat., Vol. 6(1), pp. 31-43, Feb. 1990.
- [101] H.S. Sawhney, J. Oliensis, and A.R. Hanson, *Description from image trajectories of rotational motion*, in Proc. 3rd IEEE int. conf. Comput. Vision, pp. 494-498, Dec. 1990.
- [102] D.H. Marimont, *Inferring spatial structure from feature correspondence*, Ph.D dissertation, Stanford University, Stanford, CA, Mar, 1986.
- [103] 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., pp. 576-581, Nov. 1989.
- [104] R. Hartley, *Self-Calibration from Multiple Views with a Rotating Camera* Proc. Third European Conf. Computer Vision, J.-O. Eklundh, ed., Vol. 800-801, pp. 471-478, May 1994.
- [105] G. Stein, *Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Error*, Proc. Fifth Intl Conf. Computer Vision, pp. 230-236, June 1995.
- [106] L. Lange, *Deriving the Equations of the Sections of a Cone*, The American Mathematical Monthly, Vol. 63(7), pp. 488-491, 1956.
- [107] M. Sezgin, B. Sankur, *survey over image thresholding techniques and quantitative performance evaluation*, Journal of Electron. Imaging, Vol. 13(1), pp. 146-165, 2004.
- [108] S. U. Lee, S. Y. Chung, and R. H. Park, *A comparative performance study of several global thresholding techniques for segmentation*, Graph. Models Image Process. Vol. 52, pp. 171-190, 1990.
- [109] O. D. Trier and A. K. Jain, *Goal-directed evaluation of binarization methods*, IEEE Trans. Pattern Anal. Mach. Intell. Vol. 17, pp. 1191-1201, 1995.
- [110] C. A. Glasbey, *An analysis of histogram-based thresholding algorithms*, Graph. Models Image Process. Vol. 55, pp. 532-537, 1993.
- [111] G. Taubin, *Estimation Of Planar Curves, Surfaces And Nonplanar Space Curves Defined By Implicit Equations, With Applications To Edge And Range Image Segmentation*, IEEE Trans. PAMI, Vol. 13, pp. 1115-1138, 1991.
- [112] <http://www.mathworks.com/matlabcentral/fileexchange/22683-ellipse-fit-taubin-method>
- [113] H.D Cheng, X.H, Jiang, Y. Sun and J.L. Wang, *Color image segmentation: advances and prospects*, Pattern Recognition, Vol. 34(12), pp. 2259-2281, Dec. 2001.

- [114] Toby Berk, Arie Kaufman and Lee Brownston, *A human factors study of color notation systems for computer graphics*, Communications of the ACM, Vol. 25(8), pp. 547-550, Aug. 1982.
- [115] N. Otsu, *A threshold selection method from gray level histograms*, IEEE Trans. Syst. Man Cybern. SMC-9, pp.62-66, 1979.
- [116] Kenneth R. Castleman, *Digital Image Processing*, Prentice Hall, 2nd Edition, 1995.
- [117] M. Sun, et al, *A wearable electronic system for objective dietary assessment*, Journal of the American Dietetic Association, Vol. 110(1), pp. 45-47, Jan. 2010.