

Integrated Sensors and Software for Dietary Intake Assessment

Junqing Shang

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2012

Committee:

Alexander Mamishev

Neil Abernethy

Program Authorized to Offer Degree:

Electrical Engineering

©Copyright 2012

Junqing Shang

University of Washington

Abstract

Integrated Sensors and Software for Dietary Intake Assessment

Junqing Shang

Chair of the Supervisory Committee:
Associate Professor Alexander Mamishev
Department of Electrical Engineering

Dietary intake information is valuable for understanding the relationship between diet and health, and it is becoming increasingly important as obesity rates soar. However, obtaining valid measurements of dietary intake for free-living persons is one of the most challenging problems of nutrient studies. Traditional paper-based dietary assessment methods have limitations due to bias, user burden, and cost, so improved electronic methods are needed to address important hypotheses related to diet and health. Among various approaches, image-based systems are emerging and becoming more widely used for dietary assessment with the increased popularity of cameras and algorithms. They are especially useful for food volume estimation, which is the largest source of error in calculating dietary intake.

In the thesis, the Diet Data Recorder System (DDRS) is described, where integrated sensors and software are used for the objective measurement of dietary intake for free-living persons. The DDRS has a client-server architecture, which can efficiently collect, process, and store dietary data. The client consists of a mobile sensor package and software for data collection, while the server runs the algorithms for image processing, the database, and the interface for data storage and management. The sensor package consists of a Mobile Structured Light System

(MSLS) that acquires structured light data for food volume estimation. The hardware design for the MSLS integrates a smartphone and a laser attachment. The laser attachment projects patterns onto the food, and the phone's camera acquires video clips that include the laser patterns. After presenting the entire DDRS design, data analysis that estimates the food volume is emphasized. Algorithms that extract food regions in the video, identify the laser pattern in the video frames, localize the intersection points, and generate depth maps for the corresponding images are applied. The estimated food volume, together with direct entries of food questionnaires and voice recordings, could provide dietitians with more complete food descriptions and more accurate food portion sizes. The results prove the feasibility of the proposed approach for assessing dietary intake with the DDRS. Future work will focus on automating and integrating the entire processing pipeline.

TABLE OF CONTENTS

List of Figures	iv
List of Tables	vi
Chapter 1. Introduction	1
1.1 Motivation.....	1
1.2 Scientific and Engineering Challenges	2
1.3 Scope of the Thesis	3
1.3.1 Theoretical Design	4
1.3.2 Data Analysis	5
1.3.3 Experimental Work.....	6
1.4 Contributions.....	6
1.5 Thesis Outline	7
Chapter 2. Background	8
2.1 Traditional Dietary Assessment Methods	8
2.1.1 Doubly Labeled Water Methods	8
2.1.2 Self-report Methods	9
2.1.3 Limitation to the Paper-based Food Dietary	10
2.2 Electronic-based Systems	11
2.2.1 Electronic Diaries in Health Related Research	11
2.2.2 Early Electronic Food Diary	11
2.2.3 Electronic Food Diary Based on Images	12
2.2.4 Summary of Electronic Food Diary	14
2.3 3D Reconstruction from Images	14
2.3.1 Overview of 3D reconstruction.....	14
2.3.2 Principles of triangulation-based 3D reconstruction.....	16
2.3.3 Structured light systems.....	19
2.4 Chapter Summary	20

Chapter 3. DDRS Design.....	21
3.1 Overview of DDRS Design	21
3.2 Hardware Design of Data Collection Device.....	23
3.2.1 Smartphone	24
3.2.2 Laser Module	25
3.2.3 Circuit Module	26
3.2.4 Plastic Housing	29
3.3 Software Design on the Smartphone.....	30
3.4 Design of Mobile Structured Light System	36
3.4.1 Euclidean 3D Reconstruction using a Single Integrated System.....	36
3.4.2 Hardware and Software Design for MSLS	37
3.5 Design of Database and Interface	39
3.6 Design of Volume Reconstruction.....	41
3.7 Experiments	42
3.8 Chapter Summary	44
Chapter 4. Data Analysis for Volume Estimation	45
4.1 Overview of Data Analysis and Food Volume Estimation.....	45
4.2 Food Video Segmentation.....	47
4.3 Cross Point Identification	48
4.3.1 Grid Line Extraction	48
4.3.2 Intersection Extraction	49
4.3.3 Central Point Identification.....	49
4.4 Depth Map Generation.....	50
4.5 Experiment Result.....	52
4.6 Chapter Summary	58
Chapter 5. Conclusion and Future Work.....	59
5.1 Conclusion	59
5.2 Future work.....	60
Bibliography	61

LIST OF FIGURES

Figure 1.1. <i>Thesis scope</i>	4
Figure 2.1. Principle of a pinhole camera (courtesy [61]).	17
Figure 2.2. Geometry of a pinhole camera.	18
Figure 2.3. Triangulation between a laser and a camera.....	19
Figure 3.1. DDRS overview.....	22
Figure 3.2. The integrated sensor package next to a US quarter.	23
Figure 3.3. An unfolded sensor package.....	24
Figure 3.4. Mega diffraction matrix with a point laser source.....	26
Figure 3.5. A previous laser design without mirror.	26
Figure 3.6. The schematics diagram of the circuit design.	28
Figure 3.7. A photo of the circuit module.....	29
Figure 3.8. The housing design: a) front view and b) back view.....	30
Figure 3.9. The screen shot of the main menu in the software.	32
Figure 3.10. Two screenshots in: a) start of study, and b) end of study.	33
Figure 3.11. Two screenshots of: a) meal place and b) meal type.....	34
Figure 3.12. Two screenshots of: a) Bluetooth pairing and b) video collection.	34
Figure 3.13. Three screenshots of: a) audio and text descriptions, b) choice of description time, and c) alarm for entering the description later.	35
Figure 3.14. Three screenshots in: a) “Review/Edit a Meal”, b) “Record a Forgotten Meal”, and c) “End of Day”.....	36
Figure 3.15. The distance and angle between the camera and laser.	39
Figure 3.16. The web frontend interface of the DDRS.....	41
Figure 4.1. Concept of volume reconstruction in DDRS.....	46
Figure 4.2. The flowchart for food volume estimation.	47
Figure 4.3. Workflow of the video segmentation algorithm.....	48
Figure 4.4. Valid and invalid intersections.	49

Figure 4.5. Food segmentation on an image from Internet.....	52
Figure 4.6. Food segmentation on an image collected in our lab.	52
Figure 4.7. Video segmentation on the starting frame in a video sequence.	53
Figure 4.8. Video segmentation on a middle frame in a video sequence.	53
Figure 4.9. A pair of images with and without laser grids for testing ADMR.	54
Figure 4.10. Calibration image for ADMR.....	54
Figure 4.11. Nonlinear regression of the center points.....	55
Figure 4.12. Average percentage error using regression and ADMR.....	56
Figure 4.13. Average percentage error with noises using regression and ADMR.	57
Figure 4.14. A pair of images with and without laser grids.....	57
Figure 4.15. 3D model of the reconstructed mango.....	58

LIST OF TABLES

Table 3.1. System usability test from the study participants.	43
Table 3.2. How long the participants are willing to record.	44

ACKNOWLEDGEMENTS

First of all, I would like to express my greatest gratitude to my advisor, Professor Alexander Mamishev, for his constant support and encouragement, as well as for his scientific advice and vision. I am also grateful for Professor Mamishev's constant pushing of me to produce better work, while still providing me with enough freedom to pursue my research and my career. I would also like to thank my committee member, Professor Neil Abernethy, for his valuable contributions and for his help.

I am thankful to my lab partners, Kishore Sundara-Rajan, Hsiu-Che Wang, Yuri Dvorkin, and TsrongYi Wen. It is a great pleasure to study and work with them.

Numerous experiments in this thesis were conducted by several talented undergraduates under my supervision, and I would like to thank them for their contributions. I would also like to thank the English writers in our lab, who helped me polish my papers, as well as this thesis.

I am thankful to the nutritionists at Fred Hutchinson Cancer Research Center for their support of this project.

Finally, I would like to thank my parents, Zilong Shang and Hua Zhang, and my wife, Yu Deng. Their love and encouragement have enabled me to keep pursuing my graduate study.

DEDICATION

This thesis is dedicated to my parents, Zilong Shang and Hua Zhang, and my wife, Yu Deng.

Chapter 1. INTRODUCTION

The work presented in this thesis is motivated by growing concerns regarding health problems related to diet, such as obesity and cancer. As technology improves, dietary assessment becomes more possible. In this study, an electronic system for recording and analyzing dietary intake is proposed, which can reduce user burden and provide more objective recording when compared to currently available approaches. Some of the work has been published in [1-3].

1.1 MOTIVATION

Dietary intake information is valuable for understanding the relationship between diet and health, which is becoming increasingly important due to soaring obesity rates. Over two-thirds of the adult U.S. population are either overweight or obese [4]. Overweight and obesity are risk factors for cancer, diabetes, cardiovascular disease, high blood pressure, and other health problems [5-7]. The annual number of deaths attributed to obesity is estimated to be as high as 400,000 in the United States [5-7]. There is no single cause of overweight and obesity; therefore, no single method exists for preventing or treating these conditions. Treatment may include a combination of diet, exercise, behavior modification, weight-loss drugs, or even bariatric surgery, in some cases of extreme obesity. The annual estimated medical costs of obesity were as high as \$147 billion in 2008, or almost 10% of the total medical spending for that year [8].

However, obtaining valid measurements of dietary intake for free-living persons is one of the most challenging problems in nutrient studies [9, 10]. Currently, doubly labeled water (DLW) is considered the gold standard for measuring food intake in free-living conditions, but DLW has important limitations, including its high cost and lack of real-time applicability. Alternative methods for measuring food consumption are often based on participants' reports, including interviewer-administered 24-hour dietary recalls (24HR), paper-based food diaries, and food frequency questionnaires (FFQ). Since these methods require participants to remember both the foods consumed, as well as estimate the portion size, they all have limitations due to

bias, error, participant underreporting, staff and participant burden, and cost. Studies have shown that these methods underestimate food intake by 37% or more [11]. In 24HR, foods and amounts are recalled from a participant's memory with the help of a trained interviewer, usually within a day from the diet time. For the food diaries, nutritionists mail the diaries to participants, collect the filled forms, evaluate the information for the types of errors, and contact the participants to complete the missing records. Particularly, 24HR and food diaries carry very high participant burden and high costs for data collection, making them generally unfeasible for use in large epidemiologic studies. On the other hand, although FFQs have been widely used for dietary assessment in large cohort studies due to their ease of administration and low cost, they cannot be used to measure the absolute dietary intake [12-14]. Therefore, it is desirable to develop better methods for measuring the amount of dietary intake.

If we are to obtain objective, real-time measures of dietary intake in large studies, we will need alternatives to the interviewer-administered 24HR and the paper-based, multiple-day food diary. Work is currently underway at the National Cancer Institute (NCI) to develop a computer-administered 24-hour recall, which has enormous potential to lower the costs of recall administration and analysis. This thesis focuses on developing a more accurate and less burdensome system, which could be an alternative to paper-based food diaries.

1.2 SCIENTIFIC AND ENGINEERING CHALLENGES

The research work presented in this thesis has several scientific and engineering challenges that can be classified into two categories, which as outlined below.

Theoretical Design: The design process for the integrated sensors and software relies on a good understanding of the procedure and requirements of the food diary. Since the participants need to use the system to collect their diets under free-living conditions, the device should be ultra-portable. Meanwhile, the device should also be capable of collecting complete and objective dietary information that is required for nutrient studies, especially the video data necessary for food volume estimation. On the other hand, the system should be powerful enough for nutritionists to process and store the data. Designing a system for recording dietary intake requires considerable research to optimize both the system characteristics and the data

collection procedures. The requirements of portability, estimation accuracy, and computing power make the system design sophisticated and challenging.

Data Analysis: The essential part of data analysis is to estimate the food volume from the acquired data. Among dietary information, obtaining food volume is especially important and challenging. The food amount is difficult for the participants to estimate and is the largest source of error in measuring dietary intake [15]. Vision-based approaches are proposed due to their non-contact nature, and Euclidean 3D reconstruction is required since an absolute scale should be calculated for accurate volumes. However, since image/video data are collected under free-living conditions, both the foreground and background are complicated. To address the complex shape of the foreground, video-based reconstruction is used for volume estimation. It is superior to approaches based on a single image because *a priori* shape information for the object is not assumed. While this approach could increase the accuracy of volume, it increases complexity of data analysis significantly. Furthermore, the video segmentation is also challenging due to the complex background. The entire processing needs to address various image processing algorithms, including segmentation, imaging system calibration, 3D reconstruction, so the data analysis is a sophisticated and challenging process.

1.3 SCOPE OF THE THESIS

Figure 1.1 shows the scope of this thesis. This thesis is the study of the design, implementation, and testing of integrated sensors and software for dietary intake assessment. The overall aim of this study is to develop an electronic system for recording and analyzing dietary intake in clinical and epidemiologic studies. Through the use of advanced technologies like miniaturized sensors, data recorders, and corresponding software, this thesis strives to develop a system design that can collect data on food consumption in real time and at moderate cost. In addition, the thesis aims to develop data analysis algorithms that provide nutritionists with more objective dietary information compared to paper-based food diaries. The scope of this thesis can be broken down into three major sub domains: a) theoretical design, b) data analysis, and c) experimental work.

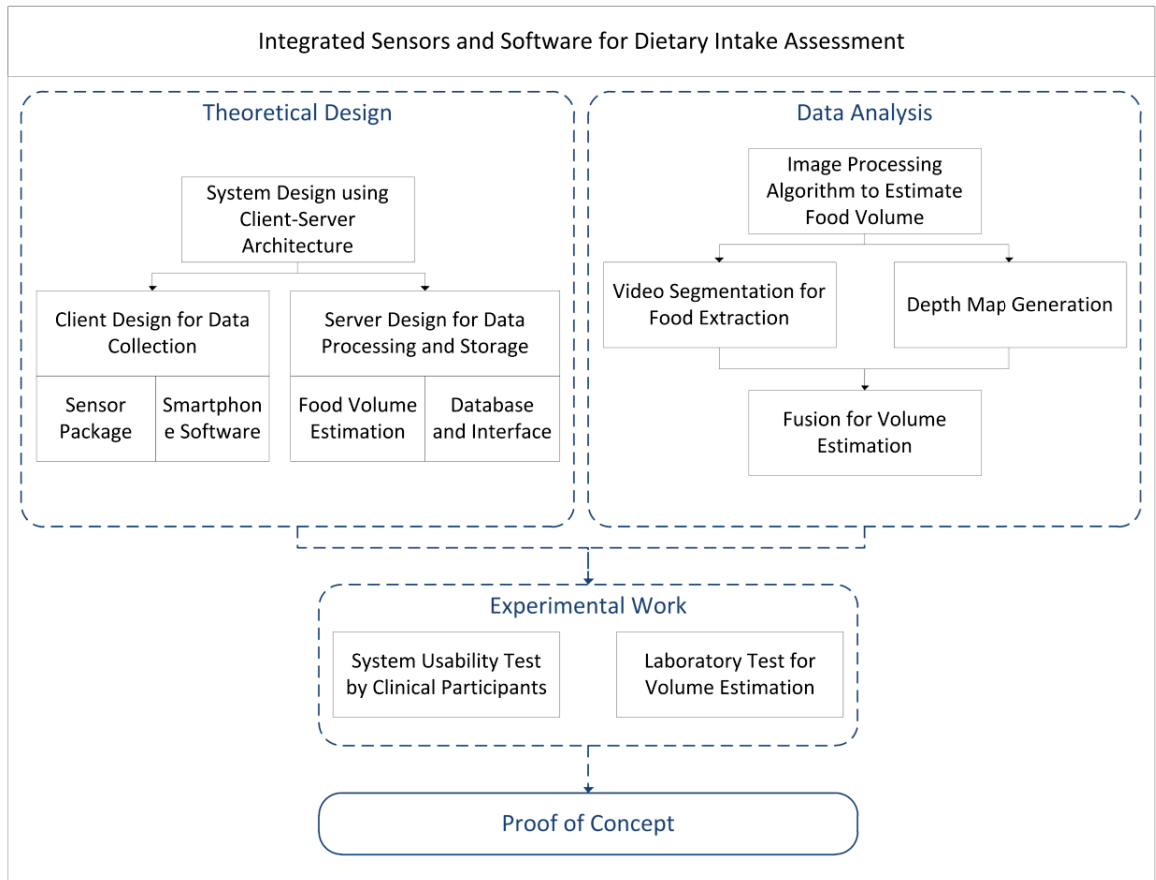


Figure 1.1. *Thesis scope.*

1.3.1 *Theoretical Design*

In the realm of theoretical design, this thesis attempts to design and develop an electronic system of food diary, which addresses the complicated scientific and engineering challenges, i.e., portability, estimation accuracy, and computing power. Since the system should be ultra-portable for participants and have enough computing power for nutritionists, a client-server configuration is designed for the system architecture. The client is designed for data collection, and the server is designed for handling the sophisticated processes of data analysis, storage, and management. In addition, a vision-based approach is designed to address food volume estimation. The system design should be sufficiently sophisticated to support routine dietary data acquisition and accurate estimation of food volume.

Client Design: The client consists of both hardware and software for the free-living participants to perform the data collection. Electronic sensors and various digital devices have been rapidly developed and quite successful in improving people's daily life during the last decade. Moreover, they have been successfully used to improve the data collection in various studies, such as physical activity data. Advances in dietary assessment can be achieved using similar sensors and devices. On the hardware side, this thesis strives to propose a sensor package for dietary data collection. The sensor package should be capable of collecting all of the data required for food diaries, and provide more objective information than the traditional method. On the software side, this thesis attempts to design and develop a software application that assists the participants and nutritionists with data collection.

Server Design: The server is designed for the nutritionists to process, store, and manage the acquired data. It consists of two major components: a) database and interface, and b) volume estimation algorithms. The database and interface are used to extract the data in a structured manner, store all the dietary information into appropriate tables, and generate consolidated reports. The volume estimation algorithms are developed to calculate the food amounts from the acquired data. This thesis attempts to design and develop these tools on the server to handle data collected by participants using the mobile client.

1.3.2 *Data Analysis*

Data analysis is crucial for converting the acquired data into nutrient information that are useful in clinical and epidemiologic studies. As aforementioned, the food amount is the largest source of error among all dietary intake measurements. This thesis attempts to develop a vision-based analyzing approach to estimate the volume from collected food video. In the proposed approach, food regions are extracted from the background using video segmentation algorithms, and depth maps are generated using 3D reconstruction techniques. The results of segmentation and 3D reconstruction are fused, leading to an estimated food volume. This thesis addresses how these image processing algorithms are applied on the dietary data.

1.3.3 *Experimental Work*

Validation is required to confirm the superiority of the proposed system over paper-based self-reporting methods, i.e., the food diary. This thesis attempts to address the validation by the following two types of experiments: a system usability test and a volume estimation test. The system usability is tested by sending the device to clinical participants, letting them acquire dietary intake information with the device for 3 days, and collecting their feedback. The volume estimation is tested by simulation and experiments in the laboratory, where the ground truth of the food volume is known. The clinical study is particularly challenging. First, the clinicians should write a rigorous study protocol. The study protocol needs to be approved by the Institutional Review Board (IRB), which may take several months. After IRB approval, it takes a huge amount of time for participant recruiting, the 3-day food diary, and analyzing the collected data.

1.4 CONTRIBUTIONS

The overall contribution of this thesis is to demonstrate the feasibility of integrated sensors and software for dietary intake assessment. Particularly, this thesis has three major contributions.

First, it proposes a theoretical design for the Dietary Data Recorder System (DDRS), which is an advanced mechanism for nutrient study compared to traditional solutions. Using sophisticated technologies could reduce participants' and nutritionists' burden, improve the accuracy, and lower the costs associated with obtaining valid measures of food intake.

Second, it presents a design for a sensor package, a Mobile Structured Light System (MSLS), which could collect active vision data for 3D reconstruction. The MSLS consists of a smartphone and a laser add-on that can be seamlessly attached to the phone. The smartphone is equipped with a video camera, accelerometer, and digital compass. The add-on contains a laser generator, a battery, and a circuit module with Bluetooth for control between the phone and the laser generator. The software on the smartphone controls the laser module via Bluetooth signals. The laser module then projects a laser grid onto the food, while the smartphone is rotated around the food to collect 360° of video. This thesis discusses many of the design considerations for the MSLS.

Third, it presents data analysis on complex dietary data that are collected under free-living conditions. To date, most image-based approaches are designed for foods of basic shapes and/or in simple backgrounds. In contrary, this thesis designs and develops image processing algorithms that could process foods of complex shapes and backgrounds.

1.5 THESIS OUTLINE

This chapter briefly overviews the motivation, scientific and engineering challenges, the scope of the thesis, and the contributions of this thesis. An outline for the thesis is listed below:

Chapter 2 reviews the related work for dietary intake assessment and 3D reconstruction. It first reviews the traditional dietary assessments, including the doubly labeled water and self-report methods. Then, the existing electronic systems are reviewed, with particular emphasis on image-based approaches for recording dietary intake. Finally, Chapter 2 also reviews the image processing algorithms that are used for data analysis in the thesis.

Chapter 3 presents the theoretical design of the DDRS. A fully-functional DDRS will include hardware and software for dietary data collection, the algorithms and protocols to process the acquired data, and the database and interface to store the DDRS information and transfer data into the clinical system used in nutrient analysis. This chapter focuses on the system architecture and describes the theoretical design of each component in the system.

Chapter 4 presents the data analysis to measure the food volume. While Chapter 3 describes the entire system, this chapter emphasizes the proposed algorithms for volume estimation. Since DDRS will be used by free-living persons, the background of the food is expected to be complex. A semi-automatic video segmentation algorithm is used to extract the food region. After that, the laser patterns are extracted from the images to calculate depth maps. The algorithms generate 3D models by fusing the depth maps from different directions. By treating the food model as a convex between adjacent 3D points, it is possible to calculate the reconstructed volume.

Chapter 5 summarizes the work in this thesis and reviews the contributions. Directions of future research will be also explored.

Chapter 2. BACKGROUND

This chapter will review previous studies on dietary intake assessment and data analysis for volume estimation. The advantages and limitations of the existing dietary assessment methods are presented, which are followed by the principles and literature review on 3D reconstruction from images. The background reviews demonstrate the significance of the DDRS, as well as the theoretical foundation and rapid research progress of the volume estimation approach based on image processing.

2.1 TRADITIONAL DIETARY ASSESSMENT METHODS

2.1.1 *Doubly Labeled Water Methods*

The traditional dietary assessments, including the doubly labeled water (DLW) and self-report methods, are well established and have been widely used in nutrient studies. The DLW method was a major breakthrough, which allowed researchers to obtain reliable and accurate estimates of habitual energy expenditure for humans under free-living conditions. The methodology using the DLW technique involves the administration of carefully weighed doses of ^{18}O and deuterium to the subject, followed by the collection of urine or saliva samples over periods of time, ranging from several days to several weeks, depending on the rate at which the water moves through the body [1-3]. This method was originally developed in the early 1950s to estimate the energy output of small animals [16]. During the first ten years of its use in humans, the DLW method was extensively validated and is now considered the "gold standard" for the measurement of total energy expenditure (TEE) of humans. Nevertheless, the DLW method is expensive, as it requires the use of sophisticated equipment to estimate the concentration of the isotope, which itself is also expensive and in short supply. In addition, since it takes at least several days to screen the water's movement through the body, DLW can neither provide clinicians with frequently accurate estimation of food intake, nor record the types of consumed foods.

2.1.2 *Self-report Methods*

The majority of popular alternatives to DLW are based on participants self-reporting, including 24-hour recall (24HR), food records, and food frequency questionnaires (FFQs). In the 24HR method, foods and amounts consumed during the previous day (or 24 hours) are recalled from memory, with the aid of an interviewer who has been trained in collecting dietary information. The primary strengths of dietary recalls are that trained interviewers can probe for details about food ingredients and preparation, and that the data collection itself cannot affect past behavior. The primary weaknesses are that participants may forget foods, report foods not eaten, or inaccurately report ingredients, food preparation methods, and portion sizes [17].

Another standard approach is the food diary or food record method. In this approach, the nutritionist design a food diary of several days, accompanying portion sizes with features, such as imbedded instructions and examples, to minimize common errors. Then, the diaries are mailed to randomly selected participants. After several weeks, nutritionists receive the filled diaries. A trained staff evaluates diaries for types of errors (e.g., missing preparation method), and analyzes records, using pre-defined coding rules to standardize data entry when food descriptions are incomplete. A second staff member then contacts the participants by phone to complete food diary documentation, and enters records for analysis. The primary strengths of the food diary are that it does not rely on memory, provided that foods are recorded in real-time and the portions can be weighed or measured. The primary weakness is that the act of keeping a diary can influence dietary behavior; therefore, not providing accurate indicators of standard behavior [18].

Both 24HR and food diary methods carry very high participant burden and have high data collection costs, making them generally unfeasible for use in large epidemiologic studies. Thus, over the past 25 years, virtually all epidemiologists conducting large studies that require dietary assessment have used list-based methods, such as the FFQ. There is a large degree of controversy in the epidemiologic and human nutrition research communities about whether FFQs are sufficiently valid to provide strong evidence about diet and disease risk [18]. It is incontrovertible, however, that FFQs cannot be used to measure the absolute intakes of energy or macronutrients [19-21], which are critical in studies of obesity or food-based exposures to environmental contaminants. Moreover, FFQs cannot capture the multiple, diverse and complex mix of cuisines consumed by many Americans [15]. There is now ample evidence to conclude

that FFQs should no longer be used in etiologic or metabolic research, and that new, cost-effective and comprehensive measures of diet are needed to advance our understanding of diet-disease relationships.

2.1.3 *Limitation to the Paper-based Food Dietary*

As mentioned in Chapter 1, this thesis focuses on developing an alternative to the paper-based food dietary, so understanding the drawbacks is crucial for understanding the motivation behind this study. The limitations to paper food diaries are well characterized and include: (a) reactivity, or a change in dietary behavior due to record keeping [22]; (b) extremely high participant burden and high required literacy, which limits use to only highly-motivated and well-educated study participants [23, 24]; (c) imprecise food descriptions, due to the fact that food composition and preparation are often unknown (especially in restaurants); (d) inaccurate portion sizes, due to difficulties in estimating portions that were not measured or weighed [25], which is a problem that is not markedly improved by providing portion size visual aids [26]; and (e) susceptibility to social desirability bias [27]. As a consequence, it is difficult to motivate study participants to maintain records over many days [28, 29]; respondents will often simplify their diets to make recording easier [18]. Additionally, respondents will be a biased sample of the underlying population of interest, portion sizes will be over- or under-estimated, defaults on preparation and composition will be frequently used when coding records for nutrient analysis, and respondents will preferentially choose and/or report foods that they perceive as socially desirable or healthy on the days that they are recording their diet. There are other practical limitations related to the administration and analysis of food diaries, which include: (1) that participants must be instructed on how to maintain a diary; (2) that completed diaries must be “documented,” which is a procedure during which trained study staff review completed records with participants to obtain missing information; and (3) that diaries must be coded and entered into nutritional analysis software, which is time consuming, complex, requires trained technicians and is expensive.

2.2 ELECTRONIC-BASED SYSTEMS

2.2.1 *Electronic Diaries in Health Related Research*

Diaries have been used extensively to measure many health-related exposures, including physical activity, sexual activity, alcohol consumption, medical symptoms, and diet. Generally, studies have found that, when assessing behavior, prospectively-collected diaries are superior to retrospective questionnaires or interviews [24]. Electronic diaries like touch-tone telephones, interactive voice response systems, and hand-held electronic devices (e.g., PDAs) have been used successfully to capture frequent, simple behavior, such as drinking [30], or subjective symptoms like pain [31, 32]. An important advantage of electronic diaries is that they can prompt participants to record behavior, either at random or at specified times [33]. More sophisticated instruments, such as electronic pill containers (which monitor medication compliance) [34] and the integration of accelerometers plus heart rate monitors (which measure physical activity) [35] have also been developed to replace or augment written records. Although the evidence is limited, studies have found that electronic diaries: (a) are more accurate because they collect real-time data and cannot be completed retrospectively [36], (b) are viewed as less burdensome by participants [37], and (c) are less subject to social desirability bias [33].

2.2.2 *Early Electronic Food Diary*

Numerous prior attempts at recording electronic food diaries have demonstrated their feasibility and potential. Electronic food diaries have been developed since the 1980's. As determining the amount of food consumed is the most challenging part of utilizing food diaries, the first approaches to electronic food diaries were based on electronic scales that recorded both food weight and description. The Portable Electronic Tape Recorded Automatic (PETRA) scale, which was developed in the 1980's by Cherlyn Electronics in Cambridge UK, records verbal descriptions and weights of foods on tape cassettes. These were used successfully by highly-motivated and well-educated persons [35], but proved to be too complex for use in low-literacy immigrant households [38]. The Nutrition Evaluation Scale System (NESSy) is a more sophisticated scale-based approach, which uses a catalog of bar-codes to identify foods placed on a scale [39]. This system is menu-driven, which simplifies its operation, but because it must be

attached to a computer, its use is limited to small-scale clinical and field studies. In summary, electronic scale methods are capable of measuring the food weight accurately under certain environments, but they have limitations for free-living persons because they are not portable and require contact with the food.

As an alternative to using scales, recent approaches to electronic food diaries have been based on direct entry of foods into computers, either alone or combined with voice recordings and food images. Voice recordings alone do not capture enough information on food composition and portion size to be useful [40, 41]. Similarly, if food images are acquired by a single camera without any fiducial markers, they can only provide qualitative information about the dietary intake. This is because a simple picture is insufficient to determine portion size. If another source of metrics is integrated into the image, food volume estimation from images becomes feasible.

2.2.3 *Electronic Food Diary Based on Images*

Numerous methods for food portion estimation using food images are emerging and becoming widely used, due to the increased popularity of digital cameras and the development of computer vision algorithms. Digital images can measure accurate portion size of foods in laboratory settings, where the camera, distance from camera to object, angle from lens to object, etc. can be rigorously standardized [42]. Image-based systems for volume estimation have already been used in the food industry, where the size and shape can be used to sort fruits and vegetables into size groups. Different methods of 3D reconstruction have been deployed to capture food volume, including stereovision [43], shape from silhouette [44], active triangulation [45], etc. Systems for measuring horticultural product size can be found in [46]. However, similar systems cannot be used for free living persons due to the portability problem.

Several small handheld devices, including CalorieKing, BalanceLog and DietMatePro, have been developed for direct entry of foods [47]. These systems simplify data collection and analysis for researchers and some incorporate date/time stamps, reminders to record meals and queries to minimize invalid entries. However, high literacy and participant motivation are required, and similar to paper records, portion sizes are frequently reported inaccurately [48]. Integrated approaches are more promising, which combine food images with food descriptions. One example is the Multimedia Diet Record (MMDR), which combines images from a 35 mm camera, food

descriptions recorded on a micro cassette recorder, and written documentation, such as recipes and food package labels [49, 50]. Unfortunately, the MMDR is burdensome to maintain and, similar to other self-report measures, total energy intake is under-reported by about 24%. In contrast, a much better example of the potential of using integrated sensor technology is the Wellnavi, a hand-held device that captures video images and digitized hand-written food descriptions, provides prompts, administers short questionnaires on food practices, and transmits data via wireless networks [51]. Using the Wellnavi is still somewhat awkward, because images must be captured at a 45° angle with a ruler-like stylus placed near the food, and users must write food descriptions on the digital screen. Registered dietitians then used the images to estimate the food consumed. Although this system did not actually perform 3D reconstruction, it did utilize the stylus as a reference.

Later, several systems based on 3D reconstruction were proposed with the development of computer vision algorithms. In [52], participants were trained to take a food image while a reference card with printed rectangles was placed near the food. Then, their system computed the food area, and assumed the linear proportionality between the area and the volume. However, this assumption is not valid under many situations, where different shapes of food, different angles for camera positions, etc. can deviate from the methodology. In [53], participants laid a card with colored grids close to the food, and took a single image by using an iPhone. The reference card was the size of a credit-card for better portability. Their algorithm used several shape templates for regular food, estimated the shape parameters from the single image and the reference card, and computed the volume from these parameters. For irregular food shapes, it used the closest template for approximation. In [54], a grid reference was also placed near the food, but multiple images were captured and stereo matching was applied. 3D reconstruction can be performed for food with rich texture, but may have limitations for textureless regions. In [55], they designed a wearable system by themselves rather than an off-the-shelf cell phone or PDA. This system also used shape templates, but required plate with a fixed size rather than a reference card. All the systems mentioned above are based on passive reconstruction methods and the scale is determined by fiducial markers, which are separated from the camera.

2.2.4 *Summary of Electronic Food Diary*

Application of electronics technology to support real-time assessment of health behavior is both practical and superior to collecting written records. With the existing electronics technology for real-time dietary assessment, it is clear that the integration of images, voice, and text for the collection of dietary intake data is now a realistic option. The availability of miniaturized sensors, ubiquitous wireless networks, and commercial software for image analysis and voice recognition makes the development of real-time electronic food diaries both feasible and timely, and the implementation of these capabilities has the potential to revolutionize the approaches to dietary assessment.

2.3 3D RECONSTRUCTION FROM IMAGES

In Section 2.2, we have presented the existing electronic food diaries, particularly the approaches based on images. In this section, we address the data analyzing algorithms. Various aspects of image processing algorithms have been involved in volume estimation, including image/video segmentation, system calibration, and 3D reconstruction. Among all image processing algorithms, this section focuses on the 3D reconstruction algorithms, since it is the essential component for food volume estimation. In the following subsections, we give an overview of 3D reconstruction in computer vision, introduce the principles of triangulation-based reconstruction, and review the structured light systems.

2.3.1 *Overview of 3D reconstruction*

Over the past few decades, 3D reconstruction from images has been an active research area in the computer vision community. Methods to reconstruct 3D models can be categorized into passive and active categories, depending on whether controlled illumination is required. Passive methods do not rely on any illumination sources; instead, they only require images taken under ambient light. Stereoscopic imaging is one of the most popular passive 3D imaging methods [56]. Simulating the human visual system, stereoscopy takes two images from different viewing angles and estimates the position of a 3D point in the scene. First, the imaging properties on the camera system are estimated using known calibration objects. Then, the 2D pixels that are projected from the same 3D point are identified in these images. After matching these 2D pixels

up, the 3D position is calculated by triangulation. Stereo matching could be further extended to more than two images, which is usually referred as multi-view stereo systems. The major challenge for stereoscopy is how to identify the corresponding 2D pixels in the images, and it is particularly difficult for the reconstruction of textureless regions.

Some alternative passive methods have been proposed to bypass the correspondence problem. Shape-from-silhouette algorithms perform 3D reconstruction by the external boundaries of the foreground objects [57]. A series of images of the same are collected by a group of cameras or a single camera but moved to different positions. A visual hull can be determined by the intersection of the viewing cones, which are defined by each camera's center of projection and the boundaries of objects. Shape-from-focus is another group of algorithms for passive 3D reconstruction, utilizing the fact that only objects close to the plane of focus appear in sharp contrast in the image [58]. As the objects can be blurred because they are either closer to or farther from the plane of focus, shape-from-focus methods usually collect multiple images from the same position, but with different focal lengths. While the preceding methods exploit only the 2D images, other passive approaches have been developed based on *a priori* information, such as the shape of the objects or the surface properties. These model-based approaches can simplify the reconstruction process, e.g., estimating the 3D model using only a single image.

On the other hand, active methods utilize illumination sources, such as a projector or laser scanner, to project a certain light pattern on the object. Many active methods attempt to generate corresponding points, similar to the stereoscopy, but by replacing one of the cameras with an illumination source. The computation required for finding correspondence is reduced, since the illumination source is controllable. Single-point laser scanning is one of the earliest active methods still being used in some industrial applications. In this method, a laser line is created and intersects with the object at a point. This single laser point is mechanically swept across the object's surface, and a camera records the motion of this point. In a pre-calibrated system, the depth can be estimated by applying triangulation calculation. Since the point needs to be moved to measure the entire surface, single-point laser scanning is a time-consuming process. Later, slit laser scanning was developed in the 1980s, and it is still popular in many commercial laser scanners. In this method, a laser projector creates a laser plane, and this plane intersects with the object at a slit. Similar to single-point scanning, this slit is then swept across the object's

surface, and a camera records the motion of this slit. The depth can be also reconstructed with known calibration information. Unlike the single-point scanning, only 1D motion is required for slit scanning, so the required scanning time is reduced significantly. By further extending the laser pattern on the object to 2D, structured light scanning is developed, which is used in the proposed system in this thesis. Literature review on structured light scanning will be addressed in Section 2.3.3. Single-point, slit, and structured light methods are all based on triangulation.

Besides triangulation methods, active 3D reconstruction also covers other schemes in a broad area. Time-of-flight is one popular approach for active imaging, where the depth is estimated from the speed of light. A single lighting pulse is emitted from the source, reflected at the object's surface, and received by a sensor. The time between sending and receiving the pulse is recorded, and thus the distance from the time-of-flight sensor and the surface can be calculated. Another kind of active imaging is model-based approaches. Simplification of reconstruction, similar to the passive model-based approaches, is achieved by assumption on the topological and optical properties on the surface. E.g., the photometric stereo, first introduced by Woodham [59], can perform reconstruction for smooth surfaces by observing the shading under three or more light sources.

Passive and active 3D reconstruction methods have evolved dramatically during the last few decades. They have provided a solid foundation for various applications in both academia and industry. More detailed surveys on image-based 3D reconstruction can be found in [60].

2.3.2 *Principles of triangulation-based 3D reconstruction*

The proposed volume estimation in this thesis is based on structured light systems, which is grounded in the concept of triangulation. Besides structured light, stereoscopy, single-point scanning and slit scanning are also based on the same concept. In this subsection, the principles of triangulation-based 3D reconstruction are introduced.

The first principle is the pinhole camera model. The pinhole model is a simple but popular geometric model for a camera or a projector. Figure 2.1 illustrates the principle of a pinhole camera model. The light rays from the objects pass through a small hole, known as the aperture, to generate an image on the file. Although modern digital cameras are usually equipped with a complex lens, they could be approximately modeled as a pinhole camera.

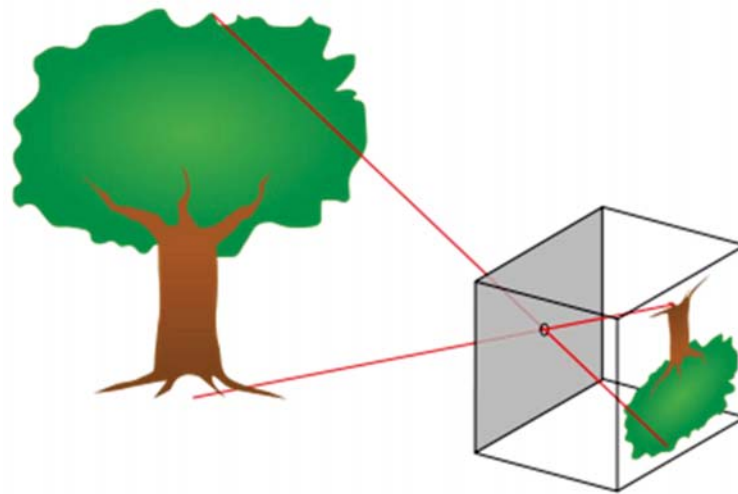


Figure 2.1. Principle of a pinhole camera (courtesy [61]).

The advantage of the pinhole model is simplifying the mathematical representation. The camera system can be represented by a plane and a point external to the plane, shown in Figure 2.2. The plane is named as the projection plane (PP), and the point is the center of projection (COP). Every point in the 3D space, except the COP itself, determines a 2D point on the projection plane. In Figure 2.2, a 3D point (x, y, z) is mapped to a 2D point (x', y') . In geometry, this 3D to 2D mapping is referred as a perspective projection. A projector can be described by a pinhole model in a similar manner.

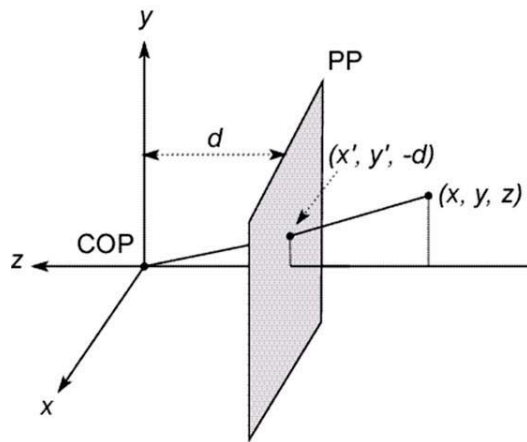


Figure 2.2. Geometry of a pinhole camera.

Using the pinhole model for the camera and the laser projector, the 3D point is linked to its 2D mappings with lines. Figure 2.3 shows an example of triangulation of a 3D point, P. In a calibrated system, the relative position between a camera and a laser is determined, so the distance from the projection center of the camera to the 3D point, $|PO|$, can be calculated by a triangular equation.

$$|PO| = \frac{|OL| \times \sin \angle PLO}{\sin \angle LPO} \quad (2.1)$$

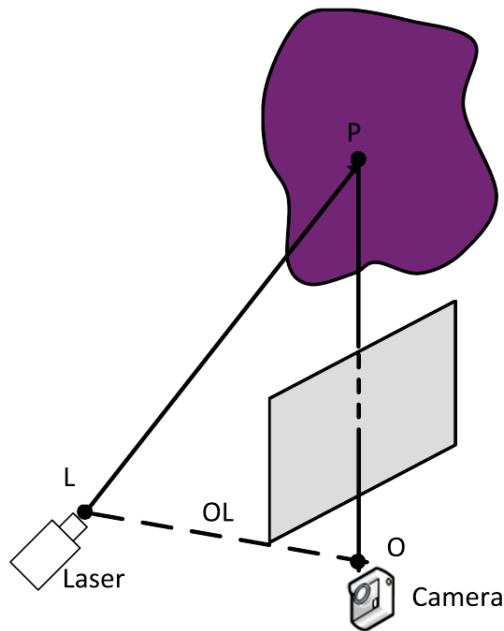


Figure 2.3. Triangulation between a laser and a camera.

2.3.3 Structured light systems

Structured light is a popular technique of active 3D reconstruction based on triangulation. It projects a pattern of light in the scene, receives the pattern deformation in the image, identifies the correspondence between the pattern and pixels, and reconstructs a 3D model of the object. Unlike single-pixel and slit scanning, structured light projects a 2D pattern onto the object, so it reduces the time required for scanning. The pattern can be coded temporally or spatially, or with a combination of both. In temporal coding, different light patterns, such as stripes with different densities, are projected onto the surface sequentially, enabling the system to find the correspondence by combining all the frames. On the contrary, the spatial coding projects a light pattern at the same time, but the pattern can be differentiated at different positions. Salvi et al. have made a comprehensive overview of structured light patterns [62].

Structured light overcomes the correspondence problem in the passive methods, so it is especially useful for textureless objects. It has been used for versatile applications, including cultural heritages [63-66], urban reconstruction [67], 3D face modeling [67], indoor scene/object reconstruction [68], etc. There have been detailed introductions and summaries on different aspects of SLS in [69, 70]. Recently, some more advanced techniques and applications

have been proposed for SLS, such as reconstruction from a single image [62, 71-76] and high-frame SLS [77]. Microsoft Kinect, while very popular in home entertainment, also use spatial structured light for depth measurement, although it utilized invisible light.

2.4 CHAPTER SUMMARY

In this chapter, we reviewed the dietary assessment and 3D reconstruction from images. Traditional methods of dietary assessment have shown severe limitations in clinical and epidemiologic studies. With technical advances, electronic approaches are emerging and becoming more and more popular for valid dietary assessments. We presented the development of the electronic systems scales and image-based food diaries. Image-based approaches show great potential, as they are non-contact and can be used under different circumstances.

Accurate measurement of image-based approaches is achieved with the development of 3D reconstruction. In this chapter, we provided an overview of the existing 3D reconstruction methods, both passive and active approaches. The principles of triangulation, the core for many scanning methods, were also introduced. Finally, the structured light was reviewed, as it is the scanning scheme in this thesis.

Chapter 3. DDRS DESIGN

Chapter 3 presents the theoretical design of the DDRS, focusing on the system architecture and the theoretical design of each component in the system. First, the overview of the system design is presented. It starts with the current data collection procedures at Fred Hutchinson Cancer Research Center (FHCRC), and explains why and how client-server architecture is designed. Then, this thesis describes the client design, consisting of both the hardware of the sensor package and the software application on the smartphone. Particularly, it explains how to build a structured light system within our DDRS. After that, the server in DDRS is tackled. The server is designed for data analyzing, storage and management. The participant's data are collected by using the mobile client, and uploaded to the server. A database with a web interface is designed to store the dietary data, and provides the nutritionists with a user-friendly interface. Using the web interface, the nutritionists can manipulate these data and generate a consolidated report for further nutritional studies. Food volume estimation is also an important part in the server. This chapter will briefly present how the food volume estimation is integrated in the system. More detailed descriptions on volume estimation are in Chapter 4, where the data analysis process is explained.

3.1 OVERVIEW OF DDRS DESIGN

The theoretical design of DDRS should address the clinical needs with the technical advances. In the paper-based food diary approach, both participants and nutritionists are involved. Participants are responsible for providing complete collection of dietary data, including filling all the required forms and returning the forms in time. Free living conditions are also important, as it would reflect participants' real dietary habit. On the other hand, the nutritionists are in charge of training on how to fill the diary information out, contacting the participants when the data are incomplete, and entering the dietary information. The nutritionists are well trained, so they can perform detailed nutritional analysis. The more complete and objective the diary data are, the more valuable dietary information could be generated. However, paper-based diaries are often incomplete and subjective, and have high user burden for both participants and nutritionists.

The DDRS is designed with a client-server configuration, aiming at a better solution for both participants and nutritionists. The client provides the participants with portability for usage under free-living conditions, and user-friendly but rigorous software for data collection. Meanwhile, it also provides nutritionists with complete and objective data, enabling good nutritional analysis. The server includes a database system with a user-friendly interface, so that the nutritionists can easily manipulate the acquired data. Also, the server is equipped with volume estimation algorithms, involving sophisticated data analyzing processes. Figure 3.1 shows an overview of the design of DDRS. First, a nutritionist at FHCRC sends a mobile client to a participant. Then, the participant carries the client for 3 days, answers the survey questions, collects the food intake, and returns the device to FHCRC. After that, the data are uploaded to the server and stored into the database. The nutritionist reviews the records and extracts useful dietary information. The food amounts are measured by applying image processing and volume estimation algorithms. The food volumes, together with other dietary information, are transferred into the Nutrition Data System for Research (NDSR), which is used to convert food items to nutrient components [78].

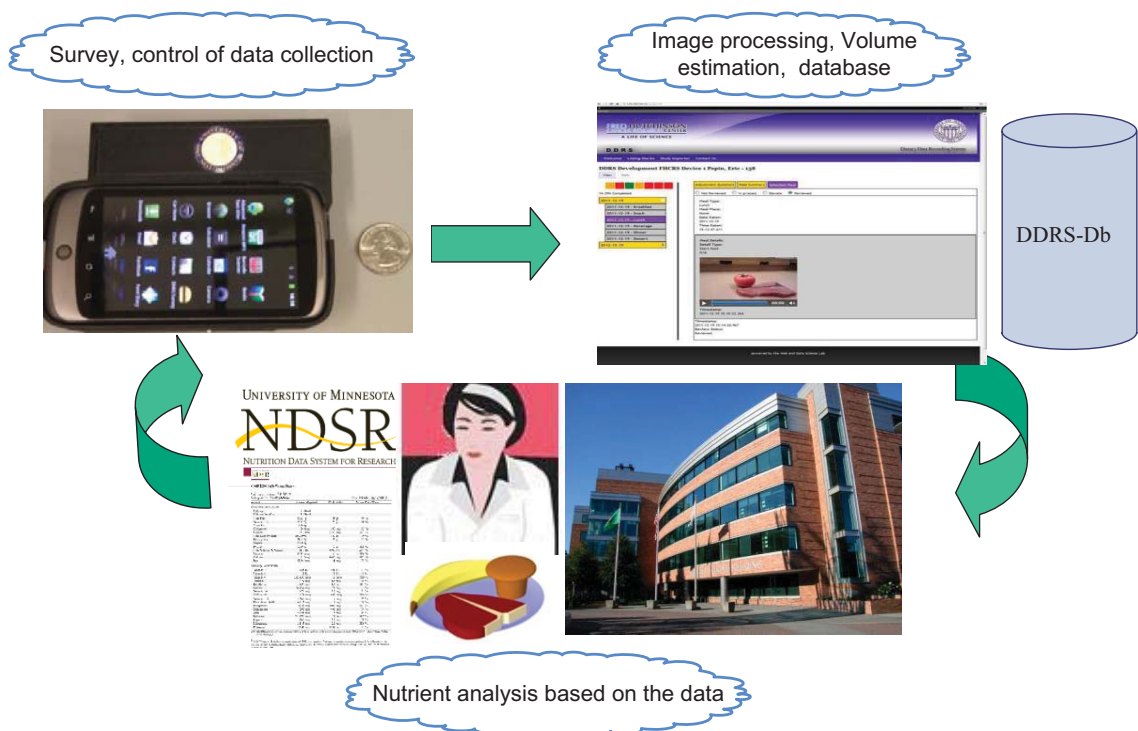


Figure 3.1. DDRS overview.

3.2 HARDWARE DESIGN OF DATA COLLECTION DEVICE

The mobile client is an essential component in DDRS. The hardware of the mobile client integrates multiple sensors into a small, hand-held device that supports complete, objective, and real-time data collection of food consumption. The design of the hardware is guided by the following goals: (1) Make the sensor package a single integrated device and keep the size of the package small enough to fit in a pocket. A single and small design is critical for portability of the system, especially under free-living conditions. (2) Reduce the burden on study participants by taking images/videos of foods, orally dictating food descriptions, and/or entering text food descriptions. This can lead to improved compliance, and reduced reactivity of diet over successive days of record keeping, and complete descriptions of foods. (3) Enable objective estimation on food portion sizes using videos. This leads to more accurate estimation of portion sizes compared to standard self-report measures. (4) Collect data in real time. This will reduce the use of “end-of-the-day” recall to describe meals and minimize effects of faulty memory on accuracy of recording food intake.

The hardware design of the mobile client consists of a mobile phone, a laser module, a battery, and a circuit module bridging the phone and laser. All of the components are mounted in a small plastic housing, so the package is portable and easy to carry for participants. Figure 3.2 shows a functional prototype of an entire mobile sensor package in customized housing. The power switches and charging ports of the phone and the laser add-on package are also illustrated in the figure.



Figure 3.2. The integrated sensor package next to a US quarter.

Figure 3.3 shows an unfolded sensor package, displaying all of the components of the hardware design. The right part of Figure 3.3 is the smartphone sitting outside the housing; the left part is the housing containing the laser module, the controlling circuit and the battery; and the top part is a plastic lid with a UW logo on it. When the package is folded, the circuit and the battery are lying under the phone inside the housing, and the laser module is covered by the plastic lid. In the following, we present all the components in the sensor package, i.e., the smartphone, the laser module, the controlling circuit, and the plastic holder.



Figure 3.3. An unfolded sensor package.

3.2.1 *Smartphone*

A smartphone is a high-end mobile phone with more advanced computing ability, larger storage size, and better programmability than common cell phones. During the last decade, the smartphone has been developed rapidly and changed the human life drastically. Current smartphones are equipped with voice recorders, digital cameras with moderate resolution, and high-resolution touchscreens that support convenient text inputs. Audios, videos, images, and text can all be easily recorded by a smartphone. In addition, many smartphones contain accelerometers, digital compasses/gyroscopes, light sensors, and GPS. All these sensors make the smartphone a suitable platform for electronic food diaries. Moreover, the smartphone supports multiple mechanisms to connect with an external PC server or the Internet, via Wi-Fi,

Bluetooth, or a USB cable. The connectivity enables easy access to the acquired data. In this thesis, Google Nexus One was chosen as the smartphone platform for further development. Nexus One integrates a good bundle of different sensors. It is equipped with a 5.0 megapixel camera, and can collect 720×480 video at 20 frames per second or higher. Besides the camera, Nexus One is also equipped with a 3-axis accelerometer and a digital compass, whose data can also be used in our reconstruction process. The voice recorder and 3.7 inch touchscreen supports audio and text inputs, while the micro USB 2.0 port, Wi-Fi 802.11b/g/n and Bluetooth v2.1 provide connectivity to other peripheral devices and the Internet. The software application on the Nexus One will be presented in Section 3.3.

3.2.2 *Laser Module*

The laser module is selected as the light source in our structured light system (SLS) since it is compact, easy to control, and of fast response to the controlling signal. Some high-end laser modules that generate structured light patterns may cost several hundred dollars, which are even more expensive than the smartphone. To reduce the cost of the entire SLS, we constructed the laser module by using a cross laser pointer, a diffraction grating from NovaLasers, and a mirror. The cross laser generates a cross green pattern that illuminates the object. The diffraction grating in our system is called Mega Matrix, which is not uniform, but has a brighter center. The cross from the laser pointer interacts with the gratings, resulting in a rectangular grid with a brighter center. The index of each line in the grid can be determined by counting how far it is from the center. Figure 3.4 illustrates a Mega diffraction matrix that is illuminated by a point laser source. A point source is shown instead of a cross source, since it can better illustrate the diffraction property. As shown in Figure 3.4, the center diffraction point is the brightest, the 5th points away from the center are brighter, and the rest points are dim.

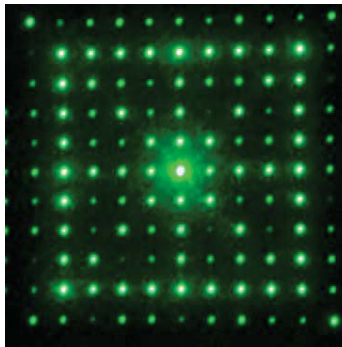


Figure 3.4. Mega diffraction matrix with a point laser source.

A mirror is a part of the laser module. The laser generator is much longer than the thickness of the mobile sensor package. If the laser pattern shoots towards the same direction as the camera directly, the laser generator will stand out of the holder. Figure 3.5 shows one of our previous designs for the sensor package. As shown, the laser generator reduces the portability of the device severely. In our current laser module, a mirror is used to reflect the direction of the laser patten, so the device is much more portable.



Figure 3.5. A previous laser design without mirror.

3.2.3 *Circuit Module*

To provide an interface between the phone and laser, the mobile client includes an electronic circuit module. Besides a hardware power switch, the module precisely turns the laser on and off in response to activation commands from the mobile application on the smartphone. Communication between the phone and circuit is accomplished through Bluetooth. The

software application utilizes the existing Bluetooth functionality of the smartphone, and a Bluetooth module is integrated into the circuit. To activate the laser in response to the Bluetooth module output, a simple switching network is utilized, characterized by the discrete MOSFET transistors and carbon film resistors. In addition, the module connects the laser to a battery power source. A 3.7 V and 850 mAh polymer lithium-ion battery is chosen to provide power to the laser and Bluetooth module. For better usability, the circuit has been designed to allow the user to easily recharge the battery power source. The recharging functionality is achieved by designing a Mini USB connector, which can be readily connected to a variety of external power sources.

Figure 3.6 shows a schematic diagram of the circuit design. The module is realized as two functional sub-circuits connected by a user-controlled hardware switch. The lower sub-circuit houses the RN-41 Class 1 Bluetooth interface and the laser generator. Upon software initialization, the RN-41 interface is paired with the appropriate smartphone. This device pairing allows for baud rates of up to 921 Kbps, although a lower rate is used in normal operation of the mobile client. To activate the laser in response to Bluetooth module output (turn on and off), a low-power switching network is utilized, characterized by Q1 and Q2. Q2 is a DMOS device that, when biased, produces a relatively high drain current, allowing the laser to be sufficiently biased and activated. This active-low switching network is primarily implemented to conserve power and extend the interval between recharge sessions.

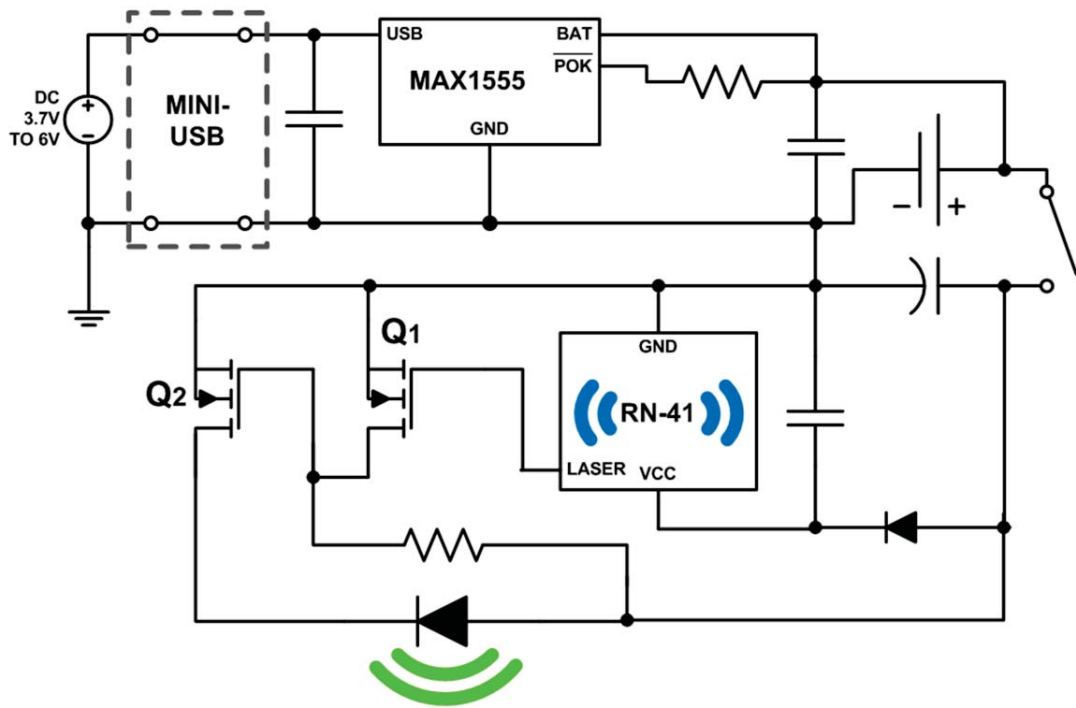


Figure 3.6. The schematics diagram of the circuit design.

The upper sub-circuit in Figure 3.6 provides the functionality of recharging the battery. Despite the efforts of power conservation, laser activation and Bluetooth communication eventually drain the battery. Furthermore, the Bluetooth module drains substantial power during idle operation. Such constant power drain can deplete the battery, resulting in the need for frequent recharge sessions. Accordingly, the function of the upper sub-circuit is to allow the user to easily recharge the battery power source. The interface for the recharging functionality is a PCB mountable Mini USB connector, which can be readily connected to a variety of external power sources, depending on the various adapters the user has access to. The MAX1555 chip functions as a USB adapter for battery recharging and while recharging the switch is kept open. When battery charge is present and the user decides to use the laser functionality of the mobile client, the switch is closed.

To extend periods between recharging, the circuit module includes a switch to turn power to the Bluetooth on or off. This switch is to be manually operated by the user, in accordance to

command prompts given through the software user interface. There are two notches for the connector and switch in the plastic housing.

To keep the module compact and durable, the electronic circuit is entirely integrated into a PCB. The small size of the module, roughly 3cm x 5cm x 1cm, allows it to be included into the system carriage without the need for any extra external design considerations. Overall, the circuit module has shown to be durable and reliable over extended testing periods. Figure 3.7 shows a detailed photo of the PCB with all the elements. The components are: 1) Class 1 Bluetooth module, 2) MOSFET laser switching network, 3) laser connection terminals, 4) a battery connector, 5) a Mini USB interface for battery recharging, and 6) a circuit power on/off switch.

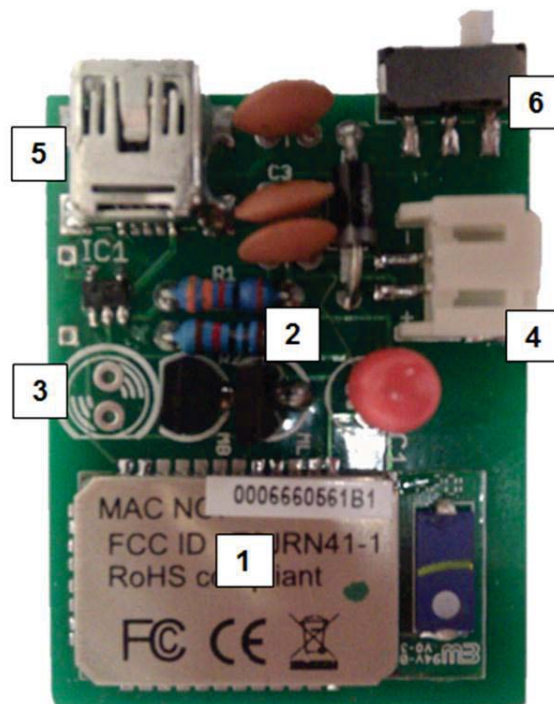


Figure 3.7. A photo of the circuit module.

3.2.4 Plastic Housing

A plastic housing is used to hold all the components of the mobile client. The housing is designed by 3D computer-aided design (CAD) software SolidWorks, and printed by a Dimension 3D printer, using black acrylonitrile butadiene styrene (ABS) plastic materials. It provides a tight fit for all components in the mobile client, so all the sensor package can be

easily delivered, carried, and used by the study participants and nutritionists. The tight fit also fixes the distance and angle between the camera and laser, enabling the precise calibration between the camera and the laser module. The housing is a little bit larger than the smartphone, and could be easily carried by the study participants. The dimension of the package is 120 mm × 85 mm × 26 mm.

Figure 3.8 shows the front and back views of the housing design. In the front view, two slots are shown for holding the mobile phone and the laser module. The circuit model is located under the phone. Moreover, the housing provides four notches for the power switches and charging ports, so the users do not need to unpack any parts during the entire study. In the back view, three holes, one square in the lower side and two circles in the upper side, are designed for light to pass through. The square hole is for the projected laser pattern, and two circular holes are for the digital camera and its flash.

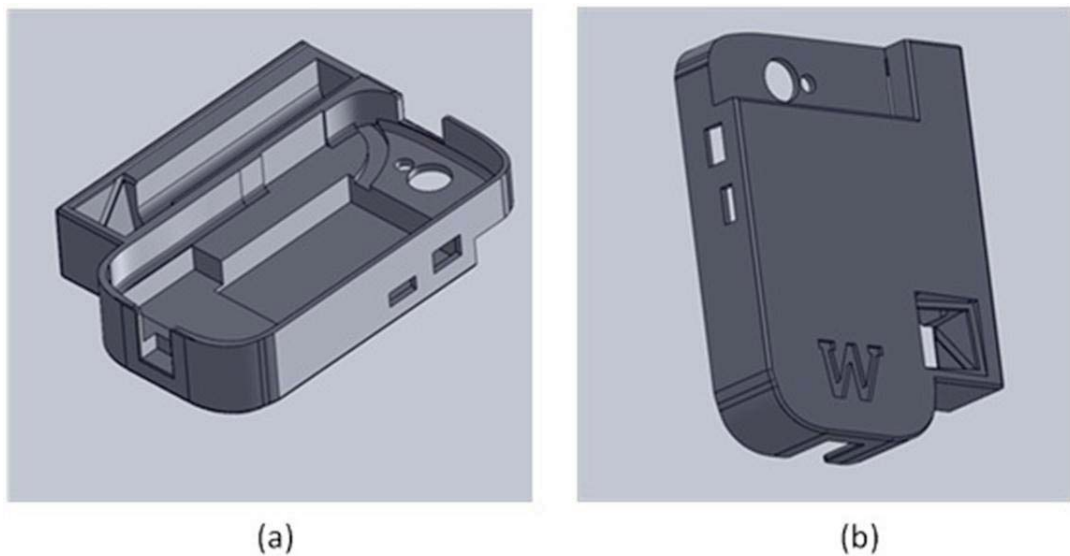


Figure 3.8. The housing design: a) front view and b) back view.

3.3 SOFTWARE DESIGN ON THE SMARTPHONE

The software design aims to support complete and objective data collection of dietary information, which is more advanced compared to the traditional paper-based food diary. There are two major functions of the smartphone software: collecting dietary data and controlling the laser module. For data collection, the software application is developed to provide a user-

friendly but rigorous protocol, so it can reduce the user burden on both study participants and nutritionists. For example, a reminder mechanism is provided in the software, reducing the chance of forgotten meals. Besides the standard questions in the paper-based method, the smartphone can collect video and audio data of the foods, which is more objective than just text descriptions of the foods. It can also collect data from a 3-axis accelerometer and a digital compass, whose data can also be used in our reconstruction process. Controlling the laser module is only used while the video is acquired. Once a study participant sets up the Bluetooth connection and starts to collect video of foods, the software will send on and off signals to flash the laser pattern. The software application is developed in JAVA on the Android operating system. Although Nexus One is chosen as the platform for development, the application could also be used on other Android-based smartphones.

The software structure and the data collection protocol are designed after multiple discussions with nutritionists at FHCRC. Figure 3.9 shows the screen shot of the main menu in the software. The “settings” button is used for general setup in the study, such as scheduling the intake days, setting up the alarms, etc. These setting are usually done concurrently when the nutritionists give the device training to the participants. The other six buttons are used by the participants during the study. They can be roughly categorized into 3 groups: 1) “Start of Study” and “End of Study”; 2) “Enter a New Meal”; and 3) “Review/Edit a Meal”, “Record a Forgotten Meal” and “End of Day”. Their functionalities are addressed in the following.



Figure 3.9. The screen shot of the main menu in the software.

Following the paper-based food diary, the software mainly collects two kinds of data, which are the answers to survey questions and the meal data. At the start and in the end of a study, a participant needs to answer a series of survey questions, mainly about his/her eating habits and summaries on the study. Figure 3.10 shows two screenshots in the “Start of Study” and “End of Study”.

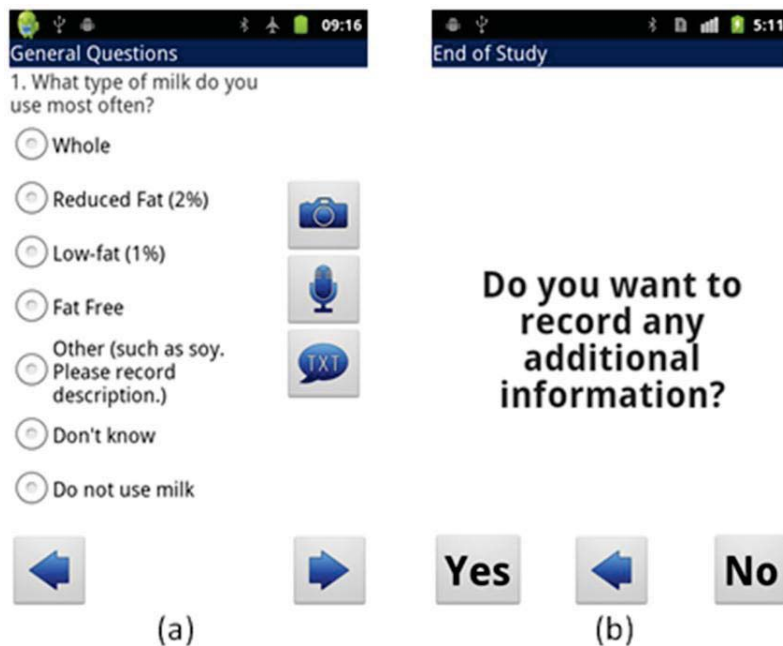


Figure 3.10. Two screenshots in: a) start of study, and b) end of study.

For the main portion of the study, the user needs to record his or her daily dietary intake. These meal data are the most important for nutritional analysis. “Enter a New Meal” enables the participant to enter all of the required information while he or she is eating. This real-time recording can reduce faulty memory and collect objective data of the foods. The study participant needs to answer questions like time and places, as well as take video of the foods. The Bluetooth on the smartphone is paired with the circuit module before the video collection. Once they are paired, the digital camera can collect video with laser patterns. Figure 3.11 shows screenshots of entering the meal place and meal type, and Figure 3.12 shows screenshots of the Bluetooth pairing and video collection.



Figure 3.11. Two screenshots of: a) meal place and b) meal type.

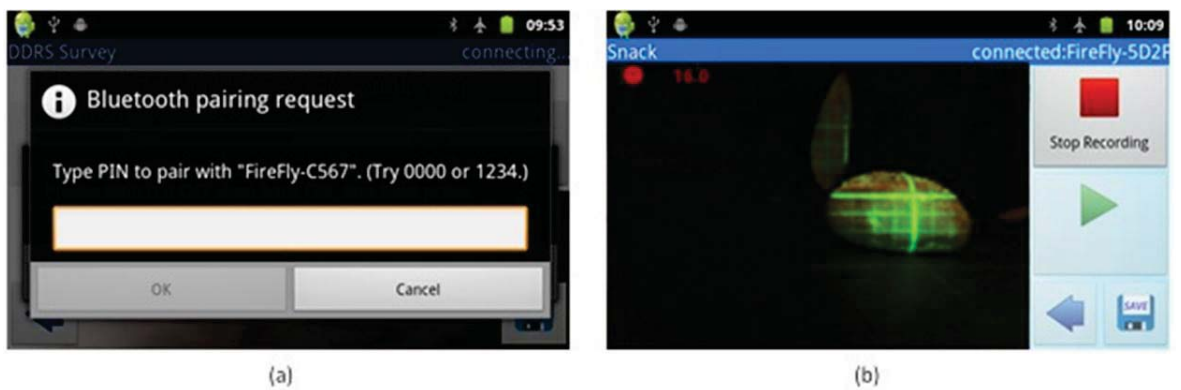


Figure 3.12. Two screenshots of: a) Bluetooth pairing and b) video collection.

After collecting the video, the participant can review the acquired video for quality evaluation, and add either audio or text description of the food, or both of them. The alarm functionality is also added to “Enter a New Meal” for a better collection process, reminding the participant to enter the description later. Figure 3.13 shows the functionality for entering the descriptions.

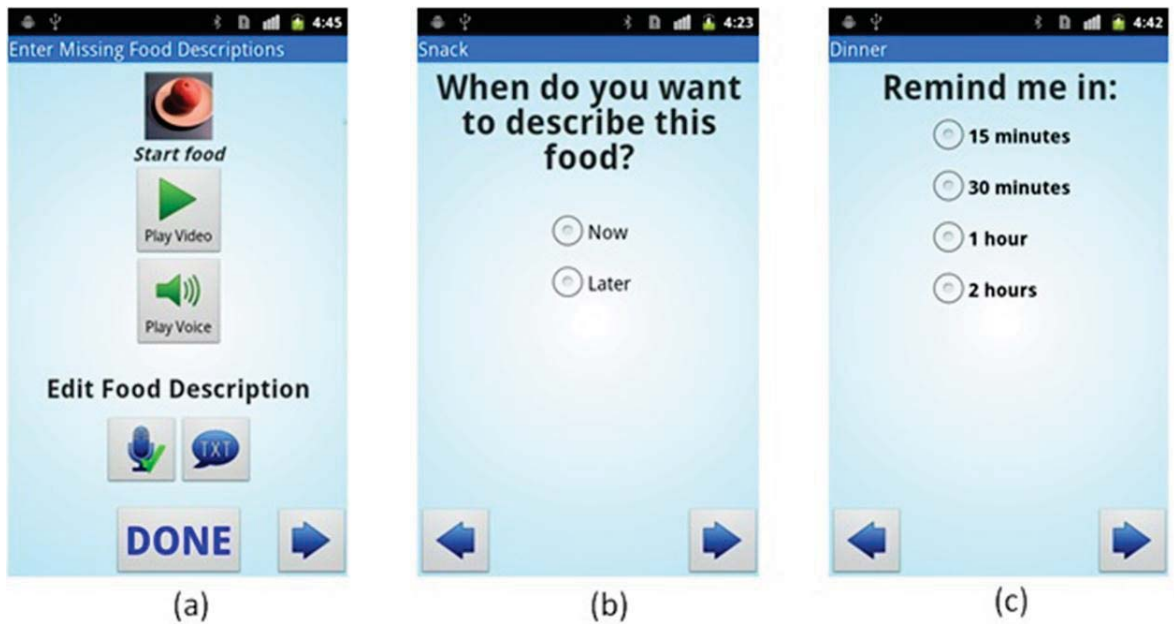


Figure 3.13. Three screenshots of: a) audio and text descriptions, b) choice of description time, and c) alarm for entering the description later.

The third group of functions is designed for reviewing and editing the foods, including “Review/Edit a Meal”, “Record a Forgotten Meal” and “End of Day”. During the entire study period, a participant is unlikely to always get things right in new meal entry. Sometimes, he or she may input an error message, or forget to record some foods. These three functions enable editing and reviewing at different times, making the overall application more robust and flexible.

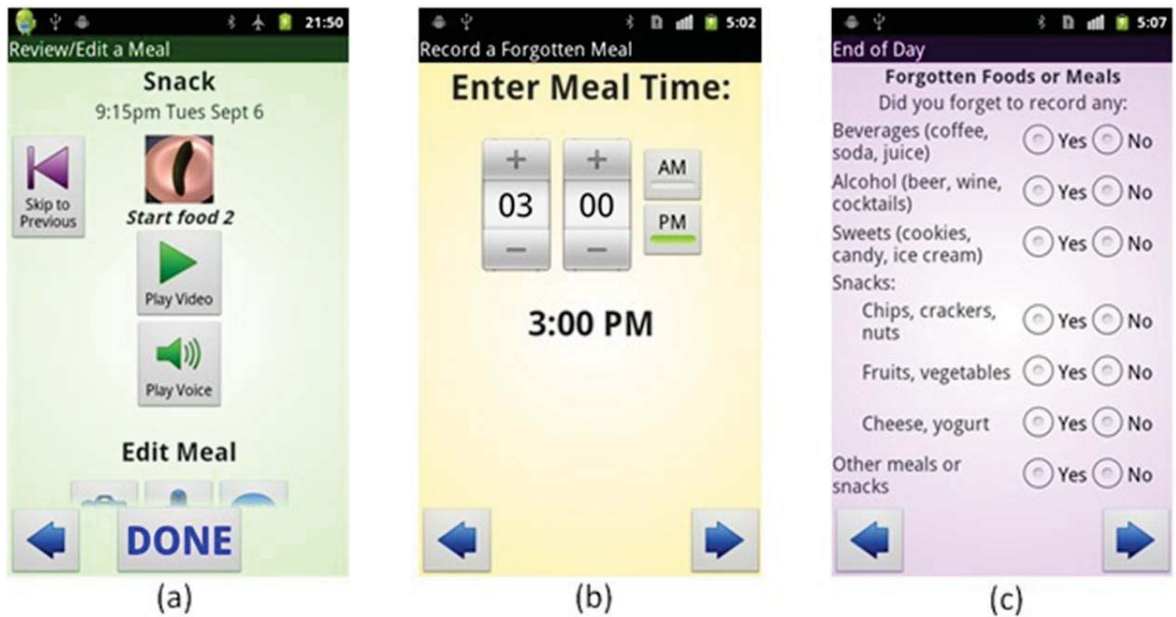


Figure 3.14. Three screenshots in: a) “Review/Edit a Meal”, b) “Record a Forgotten Meal”, and c) “End of Day”.

All the acquired data are managed by a local database, and stored into XML files. The XML structure provides a uniform data interface to the database on the server side.

3.4 DESIGN OF MOBILE STRUCTURED LIGHT SYSTEM

The design of the mobile structured light system (MSLS) involves both hardware and software components. After presenting the components in the mobile client, this section describes why and how the MSLS is designed in the following two subsections.

3.4.1 *Euclidean 3D Reconstruction using a Single Integrated System*

To measure the food intake of free-living persons, the imaging system should support Euclidean 3D reconstruction. Unlike some applications where only the shape of the object is important and metric 3D reconstruction is acceptable, Euclidean 3D reconstruction is required for food volume estimation, since the absolute scale of the reconstructed model should be known. In the 3D reconstruction pipeline of volume estimation, there must be known distances

to bring a scalable 3D space down to a 3D space with a fixed scale. In the laboratory setup, several cameras, projectors or laser scanners can be located at fixed positions, where all relative distances and angles are known. When a food item is scanned in this environment, its volume can be estimated after the 3D model is generated. However, it is not feasible to have free-living participants bring foods to the laboratory for scanning during the clinical study. Some groups have designed portable imaging systems for food volume estimation [79]. They acquire one or more images from a single camera, and perform passive reconstruction methods. All these systems reconstruct scalable 3D models by using images alone, which requires placing fiducial markers with known sizes into the scene to calculate the scale of the 3D model. These markers can be rulers, table-cloth or cards with known grids, or plates with fixed sizes. The downside of this method is that the markers are separate from the imaging system, and have to be placed into the scene by the participant. These additional markers increase the complexity of the system and raise the participants' burden.

It would be beneficial to have a single integrated system that can estimate the food volume. The MSLS, which is the essential part in the mobile client, allows for food volume estimation without any fiducial markers. The plastic housing fixes the relative position between the digital camera and the laser module, so an absolute distance is known in the system. Since the distance between the camera and the laser projector can be calibrated beforehand, it is possible to make Euclidean 3D reconstruction and estimate the food volume. We have investigated the stereo-camera system and built a prototype based on an ultra-mobile PC. However, we choose SLS over the stereo cameras because SLS is more compatible with most mobile phones in the market, requires less complex controlling and data access, and outperforms for textureless food.

3.4.2 *Hardware and Software Design for MSLS*

This section addresses how the MSLS design is optimized in both hardware and software. On the hardware side, we have already described the design of the laser module, the circuit and the plastic housing in Section 3.2. This subsection addresses optimizing the relative position between the camera and the laser. While designing a structured light system, the accuracy of depth measurement is related with the distance and the angle between the camera and the projector, the resolution of the camera, the sharpness of the projected pattern, etc. For the MSLS

design, we focus on the camera-to-laser distance and angle. According to Equation 2.1, the accuracy is proportional to the distance in a pinhole model, while the effect of the angle is much smaller. The larger the distance is, the smaller the error is caused by noises in the imaging system. This conclusion is also verified by Xu and Li In [79], where a detailed analysis is shown on how to get accurate results for SLS. In most structured lighting systems, the camera is kept far from the projector. However, the mobile client cannot tolerate a very large camera-to-laser distance due to the portability requirement. We have done simulations and experiments by changing the distance and evaluating the error of depth measurement. 70 mm is chosen as the distance between the camera and laser optical centers.

The camera-to-distance angle also needs to be carefully designed, although it does not affect the error much in a pinhole model. During the structured light scanning, the depth from the camera to food is usually not large, as the participant is usually right in front of his or her meal. A valid distance range is usually tens of centimeters. If the optical center lines of the camera and laser are parallel, 70 mm distance will make the laser pattern away from the center of collected videos. Equation 3.1 is used to calculate the angle between the camera and laser. A common depth between the camera and food is set to 40 cm, resulting in a 10° angle towards the camera. To reduce the housing size, a compact design is made by putting the laser on the bottom side of the smartphone. Figure 3.15 shows the hardware design of MSLS.

$$angle = \arcsin \frac{Dist}{Depth} \quad (3.1)$$

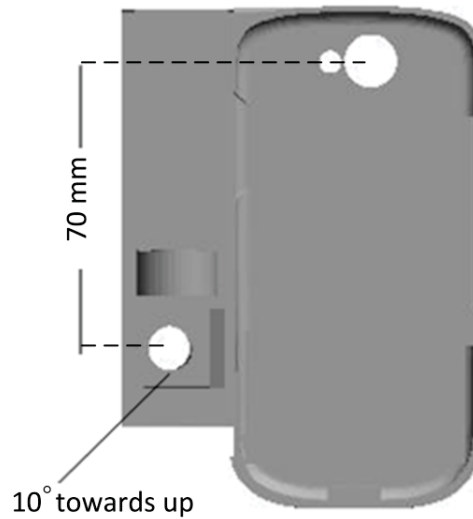


Figure 3.15. The distance and angle between the camera and laser.

On the software side of data collection, the optimization of the MSLS design is achieved by the laser flashing mechanism, which is controlled by the smartphone software. By sending out a Bluetooth signal to the circuit, the laser is flashing on and off during the video collection, resulting in video frames with and without laser grids, alternately. Image segmentation algorithms usually work better on less complex scenes, however, the projected laser patterns can generate many edge-like patterns in the video, resulting in poorer segmentation results compared to that based on videos without a laser pattern. Moreover, the frames without a laser pattern can also assist the grid extraction. During the video collection, we slowly move the camera around a food, stabilize at several positions, and collect video sequence instead of a single image to cover all sides of the food. Since the motion between two adjacent frames is small, non-grid images can be compared with according grid frames, leading to a robust grid extraction algorithm.

3.5 DESIGN OF DATABASE AND INTERFACE

A server on a PC is designed and developed for the nutritionists to process, store, and manage the acquired data from the participants. It has two major functions: a) managing the

dietary information, and b) estimating the food volume. This section focuses on the data management.

The information management system is designed using a MySQL database with a DRUPAL-powered web front-end interface. The core functionality of the system is to parse data from the mobile devices, store the data in the appropriate tables, and generate consolidated reports, as needed. As mentioned in Section 3.3, the data acquired by the mobile client are structured in an XML file format. The XML file consists of text inputs, set of images, audio files, video files, and data from accelerometers. In addition to these dietary data, the schema also stores metadata about the devices and anonymous participant information. By following the XML format, all the data are well structured and saved in the mobile device. Thus, the database can automatically parse XML files and store all these information into the server. It also includes various roles to support data input and retrieval for analytics and reporting.

A user-friendly interface is needed for nutritionists to better access the dietary information. Figure 3.16 illustrates the DRUPAL-powered web front-end interface of DDRS. DRUPAL is a popular open-source content management platform, offering the group control over the functionality of the system and reducing the amount of effort required to develop it. DRUPAL-powered sites can be configured to be viewable from mobile phones, making the output reporting system effective everywhere. The tool integrates with the Tableau-powered report generator to serve as the output reporting system to provide data to different user groups. Currently, the reporting system provides data to two different sets of users: (1) clinicians and researchers who can view aggregate data, and (2) device users (e.g., patients and study participants) who can review their own data. Clinicians and researchers will have the ability to upload, edit, and review data, while device users will have the ability to upload and review data. Both user sets will receive email notifications and mobile notifications of reports. A consolidated report is generated after the clinicians finish the reviewing process, summarizing all the dietary information. Later, these information are transferred into NDSR for further nutrient analysis.

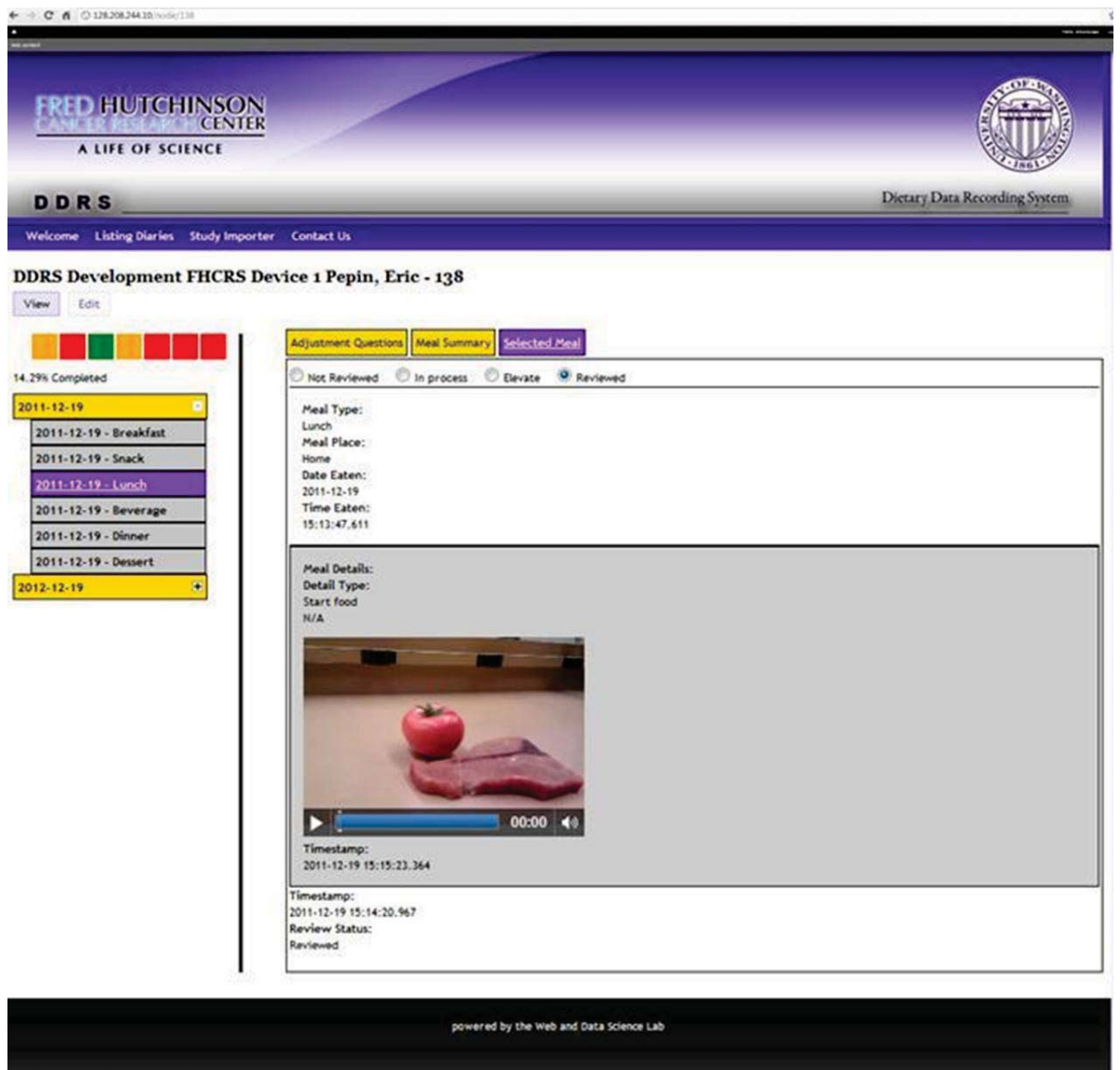


Figure 3.16. The web frontend interface of the DDRS.

3.6 DESIGN OF VOLUME RECONSTRUCTION

This section briefly introduces the design of volume reconstruction in the entire DDRS. More descriptions are in Chapter 4, where the data analysis is emphasized. Volume measurement of food is the most challenging component in DDRS. Although current smartphones are with moderate computing ability, the volume reconstruction requires extensive computation. Consequently, data analysis on volume reconstruction is performed on the server.

In addition, some human inputs may be needed for a semi-automatic algorithm, and the PC can provide a much better human computer interface than a smartphone.

With the progress of research in computer vision, many vision-based systems of measuring foods have been proposed to date. This thesis proposes an approach to measure food volume from a video camera and a laser module using active vision techniques. The laser module illuminates the food and the camera collects video sequence with the laser grid projecting onto the food. Food video segmentation is applied on the video sequence, so the food regions are distinguished from the backgrounds. Calibration is applied on between the camera and laser so that we can determine the geometric relationship between them. For a calibrated system, the algorithm for 3D reconstruction needs to identify the laser grids, generate depth maps, and register depth maps from different directions. By combining the segmentation and reconstruction results, food volume can be estimated. The reconstructed volume is important and objective information in DDRS, which can be more accurate than that from paper-based methods.

While providing data analyzing algorithms using active 3D reconstruction, the DDRS will also provide the clinician with other assisting tools for estimating the volume. Data from free-living participants can be very complex, so it is likely that the active 3D reconstruction fails under different conditions. E.g., a participant may not scan the entire food, the surface of the object may be very flaring so that it is hard to extract the laser grid pattern, or the scanning is performed outdoors where the sunshine is very strong. For such kinds of situations, an auxiliary method will be developed for volume reconstruction. The nutritionist will be able to manually mark some distances in the scene. All these tools will be integrated into the web-based interface.

3.7 EXPERIMENTS

A pilot study has been done for testing the system usability at FHCRC. We have developed three functional prototypes of the mobile client. The purpose of the pilot study is to test how the mobile client design is accepted by participants. 10 participants attended the validation study. The demographics of the participants are as follows: five men and five women, five under 40 and five 40 or older, and ninety percent highly educated (high school or equivalent n=1, college degree n=5, graduate or professional degrees n=4). Most of the participants were involved with

some level of food shopping, preparation, and meal planning. Each participant got the device and recorded a 3-day food diary using one prototype. We have developed a detailed user menu on how to use the device. Before the participants start the study, the nutritionist also gave them training on how to record data.

The overall result from the pilot study is promising, although there are suggestions on improving the client design. Table 3.1 shows the results of the system usability test from the study participants. Some participants do not test the client under all scenarios, so some rows have less than 10 feedbacks.

Table 3.1. System usability test from the study participants.

Testing scenario	Number of participants	“not at all difficult”, “not too difficult” or neutral	“Somewhat difficult” or “difficult”
Recording at home	10	9	1
Recording in a restaurant	7	5 (3 not too difficult, 2 neutral)	2 (somewhat difficult)
Recording at work	9	6	3
Taking videos	10	8 (4 not at all difficult or not too difficult, 4 neutral)	2
Recording for 3 days	10	7 (6 not at all difficult or not too difficult, 1 neutral)	3

As seen in the table, the mobile client is acceptable with most of participants in this pilot study. Besides these overall ratings, participants also commented on the system design and individual components. The overall design is user-friendly and does not require a long time for collection. At home, most of the participants accepted the data collection procedure. However, while some users were OK with using the flashing laser videos in public, others may feel uncomfortable in a restaurant setting. Some representative comments are listed from the participants:

- “It’s a user-friendly application.”
- “Really didn’t take very long to record intake.”
- “Would be reluctant to use this in a restaurant setting. I feel like it would attract a lot of attention.”
- “It was interesting recording in public; I got a lot of weird looks but that didn’t bother me.”

Table 3.2 shows how long the participants are willing to use the system for data collection. All the participants will definitely use the device for a 3-day study, while 5 of them will definitely and the other 5 will probably use it for a 5-day study.

Table 3.2. How long the participants are willing to record.

Period of days	Definitely	Maybe	No
3 days	10	0	0
5 days	5	5	0
10 or 20 days	0	5	5

3.8 CHAPTER SUMMARY

In this chapter, we have presented the theoretical design of the dietary data recorder system. The DDRS integrates sensors and software for dietary intake assessment, aiming to provide a more advanced approach than paper-based food diaries. To meet the scientific and engineering challenges of the data collection, an overall system design for the client-server architecture is proposed. After that, the thesis describes why and how each component is designed in the DDRS. The mobile client involves the design of the hardware of a sensor package and corresponding software. Particularly, a mobile structured light system is proposed, which addresses the requirements of both portability and Euclidean 3D reconstruction. The server aims to provide a centralized system for information management, as well as powerful computing power for sophisticated data analysis. Finally, the proposed design is validated by the pilot clinical study at FHCRC. The testing shows promising results for the system usability.

Chapter 4. DATA ANALYSIS FOR VOLUME ESTIMATION

Chapter 4 presents the data analysis on the acquired dietary data, focusing on the algorithms for food volume estimation. Using the mobile device, a study participant collects text inputs, a set of images, audio files, video files, and data from accelerometers. All of this dietary information is uploaded into the database and then transferred into the NDSR for an in-depth nutritional study. While other dietary information can be directly obtained from the XML files, the food amount requires sophisticated processing. This chapter presents our algorithms for data analysis.

4.1 OVERVIEW OF DATA ANALYSIS AND FOOD VOLUME ESTIMATION

Data analysis is the crucial component in the DDRS. Using the proposed mobile client, the clinicians can obtain much more complete and objective data from free-living participants. Compared to paper-based approaches, food videos with audio or text descriptions can significantly increase the quality of dietary data. The challenge then becomes how to mine useful dietary information from the objective data. For each food item, two questions need to be addressed: what is the food, and what is the quantity. Since there are a huge amount of foods under free-living conditions, it is extremely hard, if not impossible, to identify the food automatically. However, a nutritionist can review the food video and compare it with the audio and video descriptions. Usually, a well-trained nutritionist can identify foods well, in a minimal amount of time. Consequently, the food identification is performed by nutritionists based on the acquired data.

On the contrary, it is much more difficult for humans to correctly identify the quantity of food. This is why the food amount is the largest source of error in measuring the dietary intake [15]. Thus, our data analyzing algorithms focus on the volume estimation, aiming to obtain objective measurements of the food amount. Figure 4.1 depicts the concept of volume reconstruction in the DDRS. The device is rotated around the food slowly, the laser module is switched on and off rapidly, the phone collects videos, and the volume is reconstructed. Video is used for volume estimation instead of a single image from one direction since *a priori* shape information for the object is not assumed.

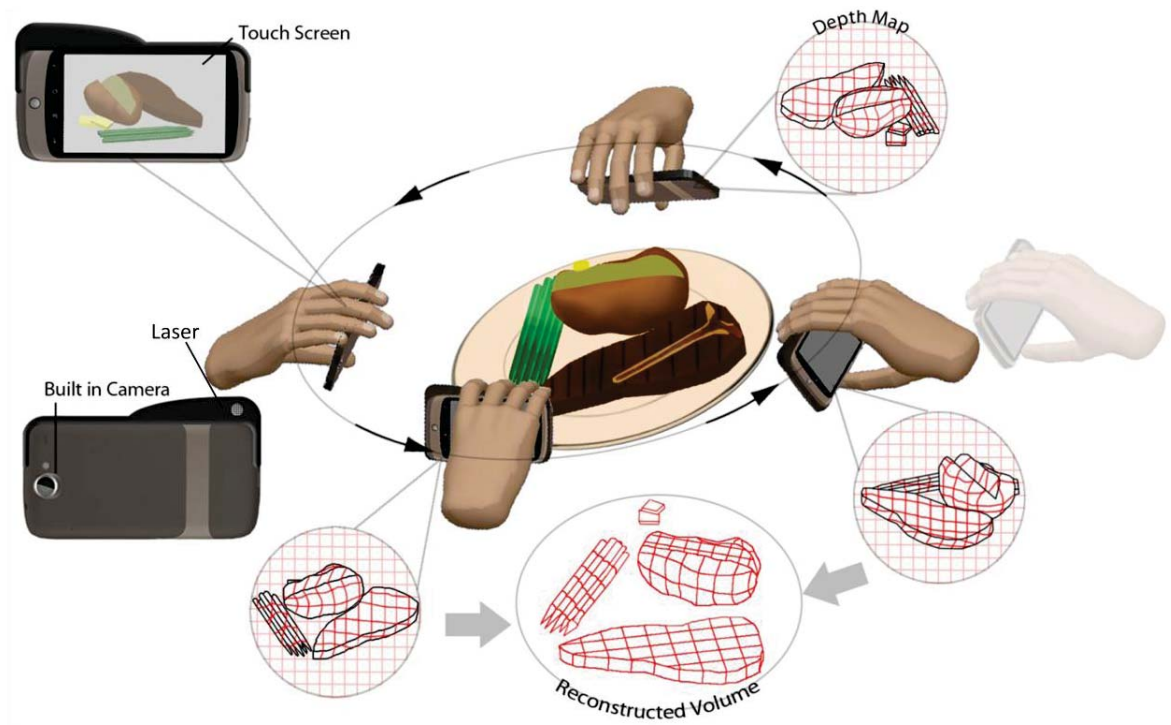


Figure 4.1. Concept of volume reconstruction in DDRS.

Figure 4.2 shows the flowchart of the algorithms used for food volume estimation. The input data are food videos with projected laser patterns. The left half of the algorithms contains the major blocks used for video segmentation. As mentioned in Chapter 3, a video sequence is extracted into pairs of consecutive frames, one image with the projected laser patterns and one without. The frames without laser are used for image segmentation. The initial segmentation is achieved by using mean-shift algorithm, implemented by the Edge Detection and Image SegmentatiON (EDISON) [3]. Then, the user draws strokes on the initial segmentation result, which is used for input of video segmentation. Next, a merging is applied across the video frames. The merging uses similar strategy as Maximal Similarity based Region Merging (MSRM) [80], but extends it from a single image to multiple video frames.

The right half of the algorithms is used for depth image generation. Before the participants collect videos, system calibration is performed in advance. Then, the acquired video frames with the laser are used to identify the laser cross. The identified crosses, together with the calibration data, are used to generate the depth maps. Finally, the results from segmentation and

depth generation are fused to calculate the 3D food model. Currently, the fusion involves manual editing, and an automatic algorithm is under development. Once the food model is created, all the 3D points of the food model are linked and the volume can be estimated. The next sections will describe each block of the algorithms, except the fusion, which will be covered in Section **Error! Reference source not found.**

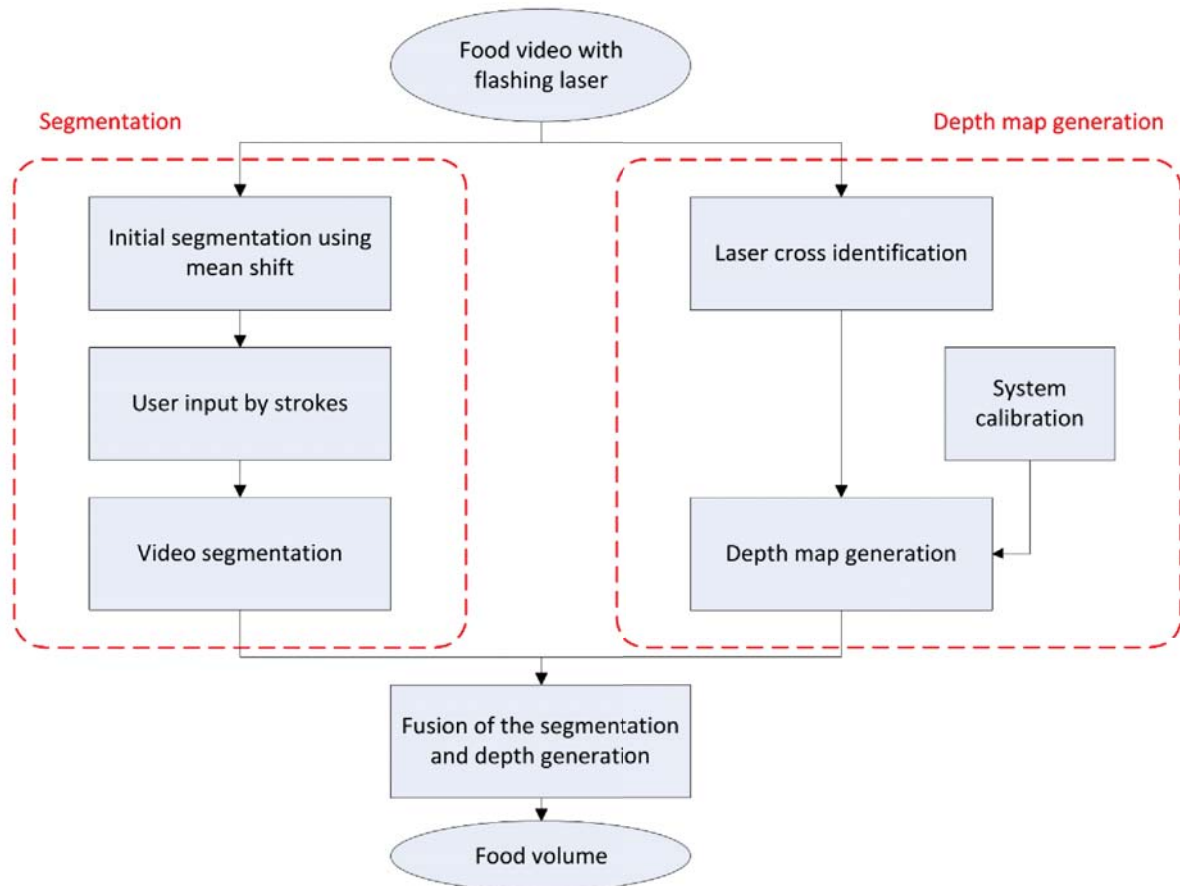


Figure 4.2. The flowchart for food volume estimation.

4.2 FOOD VIDEO SEGMENTATION

Extracting dynamic objects from video is one of the most challenging topics in image processing field. It is more difficult if the processed video is taken by the digital camera in a smartphone, which is usually in low quality, with non-stable backgrounds. In our system, the food region needs to be extracted under free-living conditions, so the background is even more complex and unstable than other cell phone applications. Non-grid video frames are chosen for

applying the segmentation algorithm, as the laser grid can destroy the boundary information in an image.

Successful image segmentation can be achieved by combination of EDISON [3] and MSRM [80], even if the background is very complicated. However, the MSRM requires user input for every input image, resulting a huge amount of work when there are many video frames. We have developed a seed selection algorithm, which will select the seed regions based on the previous picture's segmentation result and the initial segmented information of the current picture. Therefore, the user could just simply choose a starting frame and input some strokes for sampling, and all following images will be executed automatically, without requiring further user inputs. Figure 4.3 shows a detailed workflow of the video segmentation algorithm.

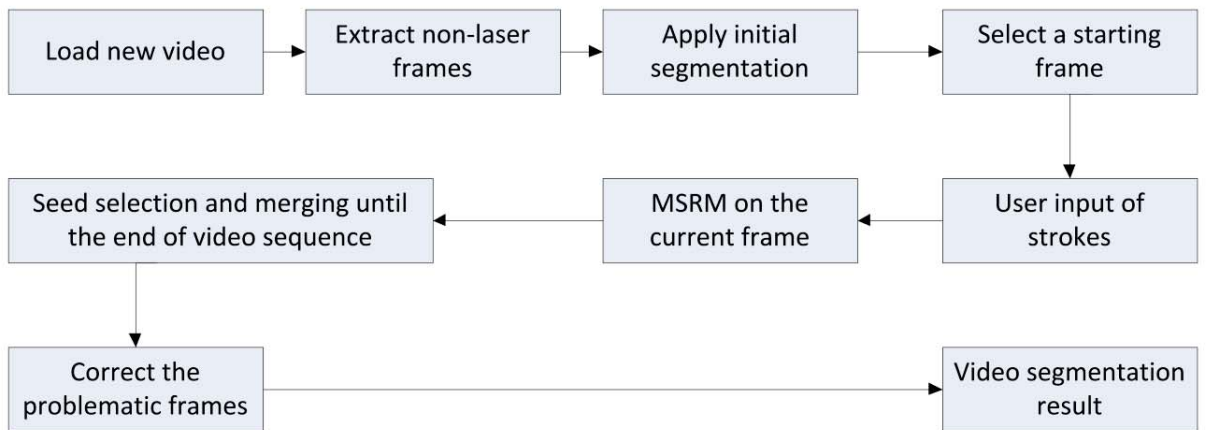


Figure 4.3. Workflow of the video segmentation algorithm.

4.3 CROSS POINT IDENTIFICATION

4.3.1 Grid Line Extraction

The laser grid lines should be extracted from the image pair for the following steps in our algorithm. As the motion between two images is very small, the non-grid image is subtracted from the grid one. After transforming the subtraction image into grayscale, a Gaussian filter is applied and subtracted from the original grayscale image. The result then is transformed into a binary image, where morphological processes can be applied. For our algorithm, isolated pixels are removed, a horizontal and vertical structuring element is used to perform morphological

closing, and then the lines are thinned to form a skeleton. Finally, the grid skeleton is stored for later processing. Assuming the two frames are I_{on} and I_{off} , the grid line extraction can be summarized by the following equation:

$$I_{skeleton} = skeleton(Gaussian(rgb2gray(I_{on} - I_{off}))) \quad (4.1)$$

4.3.2 Intersection Extraction

To generate a depth map, the intersection points in the grids need to be extracted from the grid. First, the intersection candidates are located by finding the branch points in the grid skeleton. Using the candidate points obtained, we further filter the results by obtaining the perimeter of a window around each candidate from the skeleton image. The window size can be determined by roughly estimating the density of the candidates. If the window perimeter cuts across the grid skeleton at 4 spots, the candidate is considered to be an actual intersection [81]. Figure 4.4 shows a comparison between valid and invalid intersections although both of them have local branches. The temporary results are further filtered by checking the Euclidean distance of each intersection to another. If the distance is less than a certain threshold, the two intersection locations are combined and averaged.



Figure 4.4. Valid and invalid intersections.

4.3.3 Central Point Identification

The central point in a grid needs to be identified since it is the reference for other intersection points. Central lines are much brighter than the other ones in the video frames, so the grayscale intensity is used to locate the central point. Using the similar process as intersection extraction, the perimeter of a window can be created around each intersection. The perimeter will cut across the grid skeleton at four positions. The central point has high

intensities at all four positions, other intersection points on the central lines have high intensities at only two positions and low intensities at the other two, and points off central lines have low-intensity at all four positions. After this process, the central point of a laser grid is differentiated from others.

4.4 DEPTH MAP GENERATION

With intersections extracted and the central point identified, a depth map can be created from an image pair. In a calibrated system where distances and angles regarding the camera and laser are all known, depth at each pixel can be calculated through triangulation. The requirement of our SLS is that the central point is projected and detected on the food, and the depth maps can be created even if some intersections are missing. To achieve this, we proposed an algorithm named Adaptive Distance Measurement by Regression (ADMR), which is inspired by the Splitting Cell Algorithm (SCA) [83, 84].

There are two major geometric differences between our MSLS and traditional SLS. Firstly, in MSLS, the distance between the camera and the laser is small. In our MSLS, a camera and a projector are 70 mm away from each other, while they are tens of centimeters away in a traditional SLS. Secondly, the valid working range of the MSLS is large, since it is designed for free-living people, where no rigorous valid distance can be set. In the SCA, a calibration image is segmented into rectangular regions. Each rectangular region corresponds to one line segment, and there is no overlap between adjacent rectangles to prevent mapping ambiguity. As a result, the valid range of depth is limited.

ADMR is proposed to meet the demands of our application. It contains two sub-processes: calibration and depth map generation. In calibration, we acquire a series of images within a valid range of depths from the camera to an image plane. The intersection points in all planar images are displayed in a single calibration image, where line segments are formed. Each line segment corresponds to one intersection and illustrates how this intersection moves as the depth from the image to a surface at the pixel location changes. We define the row number as M , the column number as N , and the calibration frame number as K . A calibration point at the m^{th} row, n^{th} column, and k^{th} frame is $p_{m,n,k}$. For points with the same row and column index but in different frames, they should all be in the same line segment $L_{m,n}$, which is defined as:

$$L_{m,n} : y_{m,n,k} = A_{m,n}x_{m,n,k} + B_{m,n} \quad (4.2)$$

When noise is present in the image, the points are not perfectly aligned. As a result, 1D linear regression is applied to find the best line that fits all of the points.

$$\min_{A_{m,n}, B_{m,n}} \left(\sum_{k=1}^K d^2(p_{m,n,k}, L_{m,n}) \right) \quad (4.3)$$

Once the line segments are determined, the z distance from $L_{m,n,c}$ is calculated. The relationship between x and z is:

$$z = \frac{G_{m,n}}{x + H_{m,n}} \quad (4.4)$$

A nonlinear regression is needed to estimate the parameters $G_{m,n}$ and $H_{m,n}$. We chose the Levenburg-Marquardt algorithm to perform the nonlinear least squares regression.

Once the calibration is finished, four parameters are estimated for each line segment. They are $A_{m,n}$ and $B_{m,n}$ estimated by linear regression, and $G_{m,n}$ and $H_{m,n}$ estimated by non-linear regression. This calibration model represents all information for a fixed SLS, so we do not need to calculate the distances and angles between the camera and laser beams.

The second sub-process is the depth map calculation. For an incoming cross point, we find the line segment that is closest to this point, and then identify its real depth. However, to support a large working range suitable for free-living conditions, the entire line segments are very close to each other. To increase the robustness with noisy image with noise, an adaptive range selection is applied. We assume that the central point of the laser grid is always on the food items and that the common foods are not of a very large volume. When we estimate the depths of the points, only those close to the central laser points are used. If the entire valid ranges of the calibration data are R_{min} and R_{max} , the center point depth is d_c , and the maximum food size is s . We can define the min and max range for the valid depth estimation as:

$$r_{min} = \max(R_{min}, d_c - s) \quad (4.5)$$

$$r_{max} = \min(R_{max}, d_c + s) \quad (4.6)$$

Using r_{min} and r_{max} , we can find the corresponding x_{min} and x_{max} of a line segment. A fragment of the $L_{m,n}$, $L_{m,n,c}$, is defined as:

$$L_{m,n,c} : y = A_{m,n}x + B_{m,n}, \text{ where } x_{min} \leq x \leq x_{max} \quad (4.7)$$

After depth maps are created, they can be registered to generate a 3D model. Currently, the fusion involves manual editing.

4.5 EXPERIMENT RESULT

In this section we will show the experiments of our data analysis. Both the segmentation and 3D reconstruction results are presented.

We start testing with the image segmentation result. Both images from the Internet and images acquired in our lab are tested. Figure 4.5 shows how a sandwich is segmented from its complex background by some simple user inputs, and Figure 4.6 shows how a muffin that collected in our lab setting is segmented from a dark background.

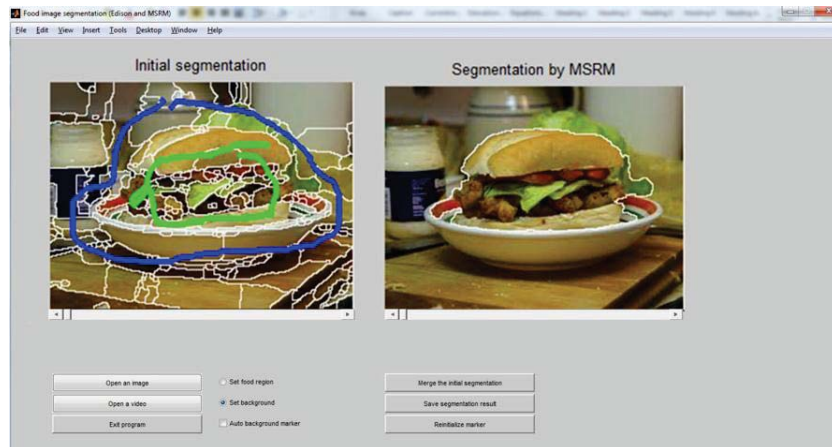


Figure 4.5. Food segmentation on an image from Internet.

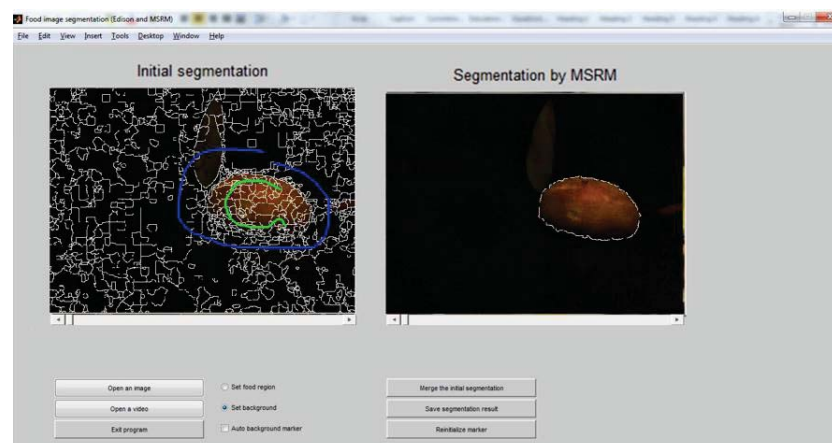


Figure 4.6. Food segmentation on an image collected in our lab.

Segmentation is performed across the video sequences. Figure 4.7 and Figure 4.8 show the segmentation result of a food video sequence, a tomato in a plate. The stroke are marked in the starting frame, and the boundary is outlined in Figure 4.7. The rest of the sequence is automatically marked and the segmentation is applied. Figure 4.8 shows the segmentation result of a frame in the middle of the video sequence.

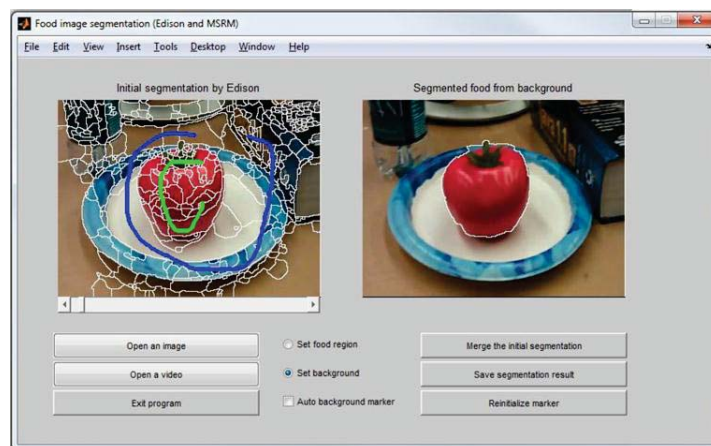


Figure 4.7. Video segmentation on the starting frame in a video sequence.

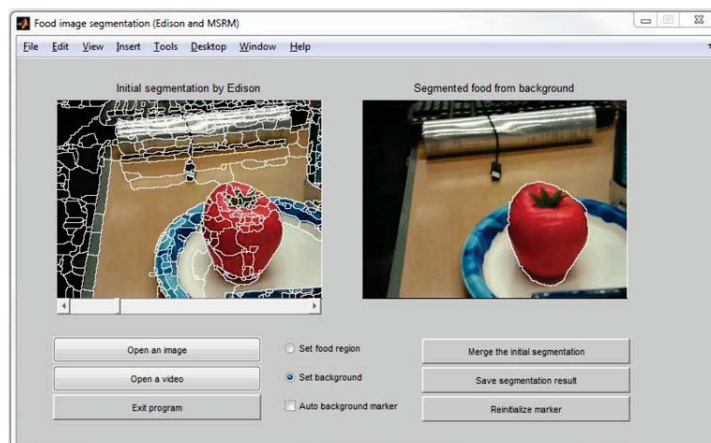


Figure 4.8. Video segmentation on a middle frame in a video sequence.

To evaluate the ADMR, we collected 13 parallel planar images that are perpendicular to the center of the camera. These images covered a range of 300 mm and one of them is shown in Figure 4.9.

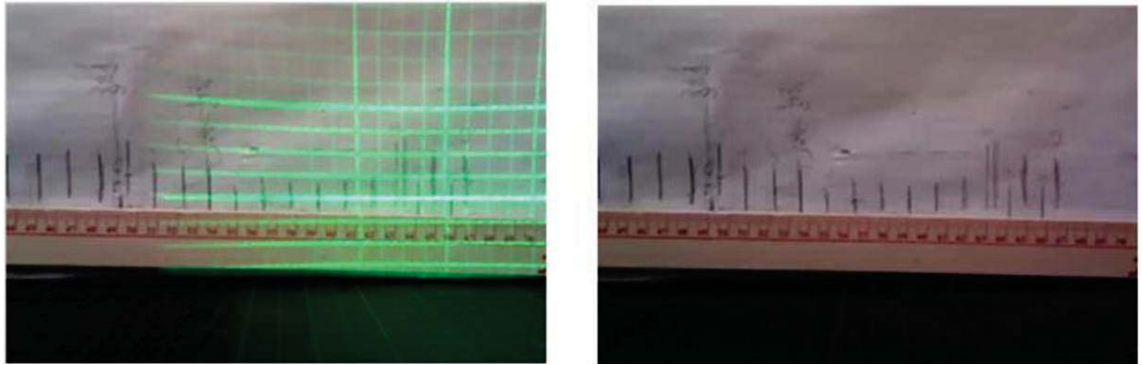


Figure 4.9. A pair of images with and without laser grids for testing ADMR.

A calibration image can be generated from these planar images. Figure 4.10 shows the line segments generated by ADMR. Each line segment represents one laser point while the distance changes. The model parameters ($A_{m,n}$, $B_{m,n}$, $G_{m,n}$, and $H_{m,n}$) are estimated from the calibration data. As we can see, there are significant overlaps in the calibration data, so SCA will fail in this setting.

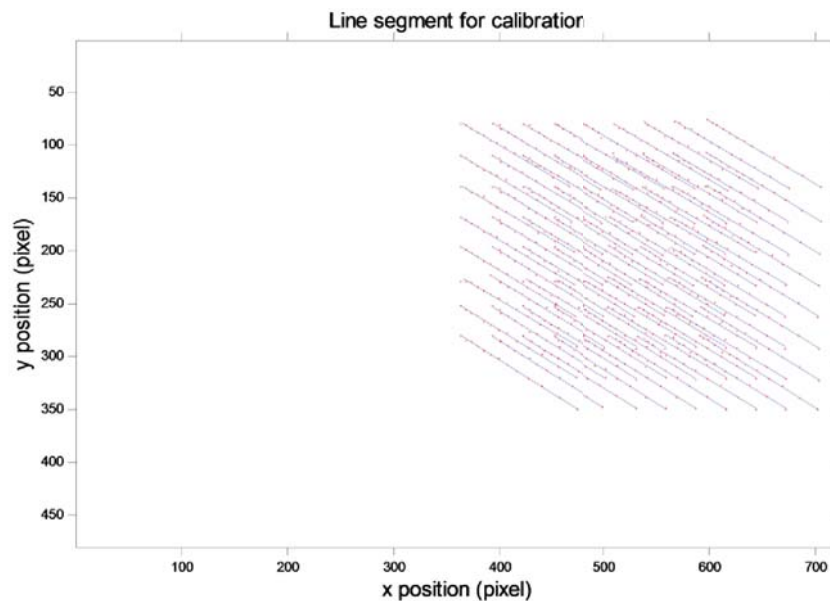


Figure 4.10. Calibration image for ADMR.

In ADMR, the center point is used as the reference. Figure 4.11 shows the nonlinear regression result of the center points. It fits Eq. 4.4 well.

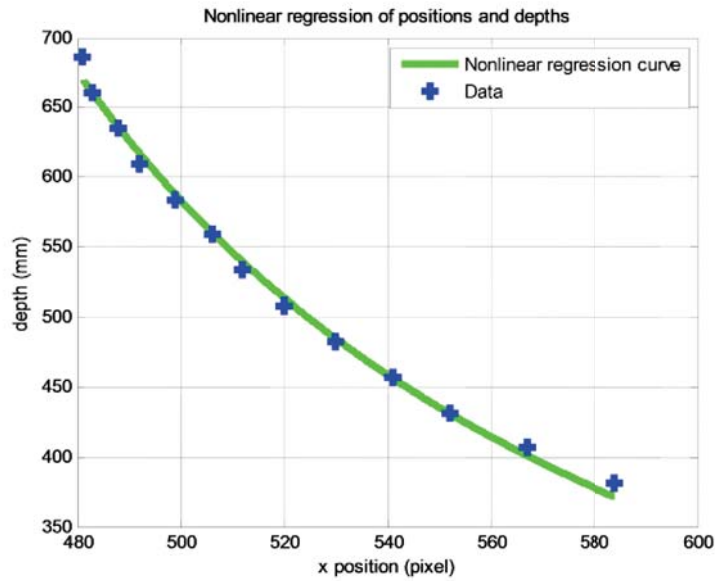


Figure 4.11. Nonlinear regression of the center points.

Using these images, we tested the performance of the ADMR. Each time, we selected one out of 13 images and used the remaining 12 frames as the calibration image. The 1st and 13th images were not selected for testing as they are at the end of the range, and the other 11 images, from the 2nd to the 12th images, are all tested for the accuracy evaluation. For comparison, we also generated the result that only uses regression. We evaluated the errors between the estimated distances and actual distances, as shown in Figure 4.12. When a point was mismatched to another line segment, we counted the number of mismatched points. For the regression method, 3-point pairs were mismatched, while there was zero mismatch in the ADMR.

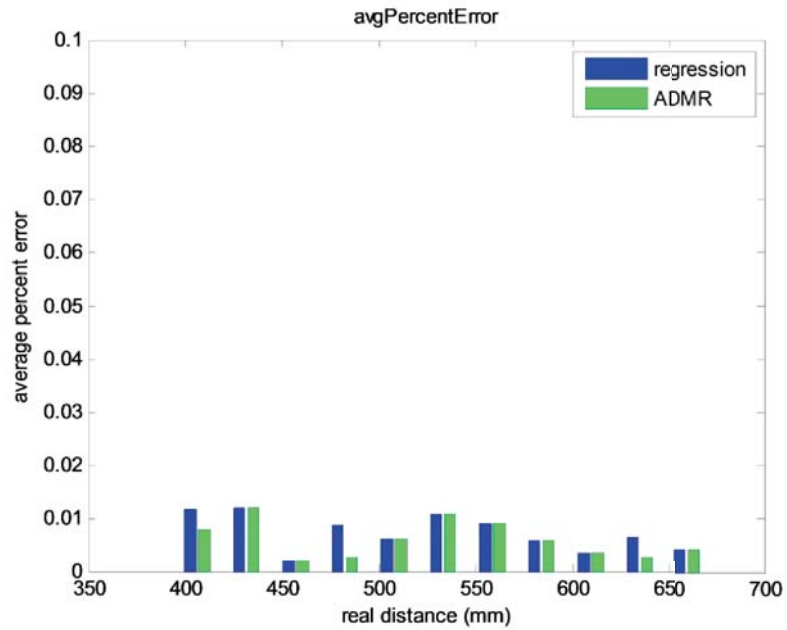


Figure 4.12. Average percentage error using regression and ADMR.

When noise is present in the input image, aSCA could perform much better than SCA. We added normally distributed noise ($mean=0$, $std=1$) to the x and y position of the laser grids, and the average error is shown in Figure 4.13. There were 32 mismatched pairs in the regression method, while zero mismatch occurs in the ADMR.

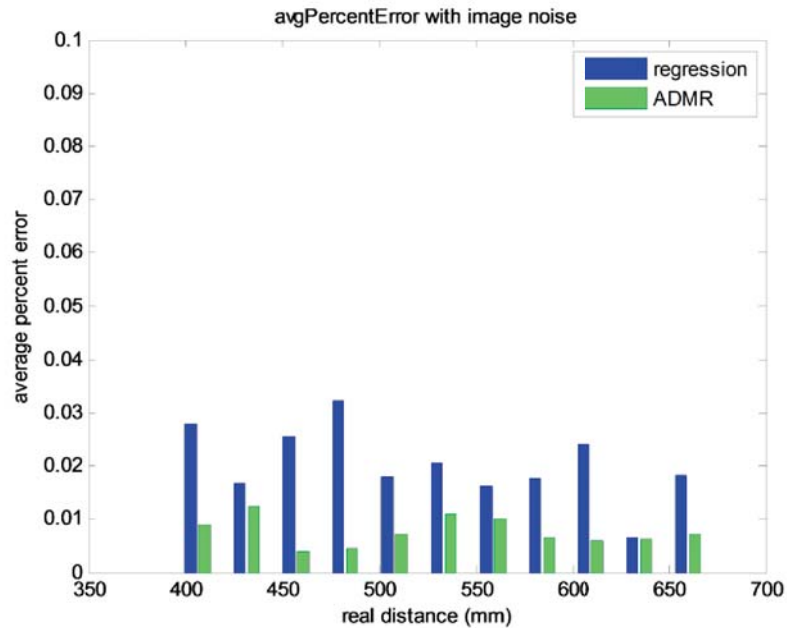


Figure 4.13. Average percentage error with noises using regression and ADMR.

Using the proposed ADMR, we can generate depth maps for foods. Figure 4.14 shows a pair of images with and without grids on a mango. The grid quality is not perfect since the laser cross from the generator is not well focused. A better laser generator may improve the laser quality.

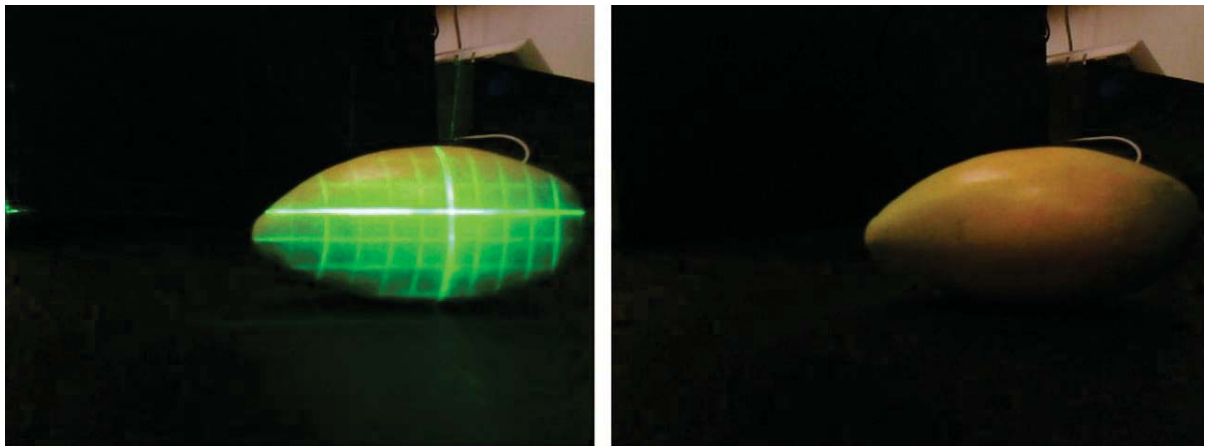


Figure 4.14. A pair of images with and without laser grids.

Depth maps can be created with the proposed ADMR. Once the depth maps from all directions are created, they can be registered for a 3D model. Google Nexus One is equipped with a digital compass, so we can get the rotation of each depth map. Currently we apply a

manual translation based on the image features. Figure 4.15 is the 3D model of the reconstructed mango by using six depth maps. The actual volume measured by a water tank is 260 cm^3 , while the estimated volume is 280 cm^3 . The error of this estimate is 8.3%.

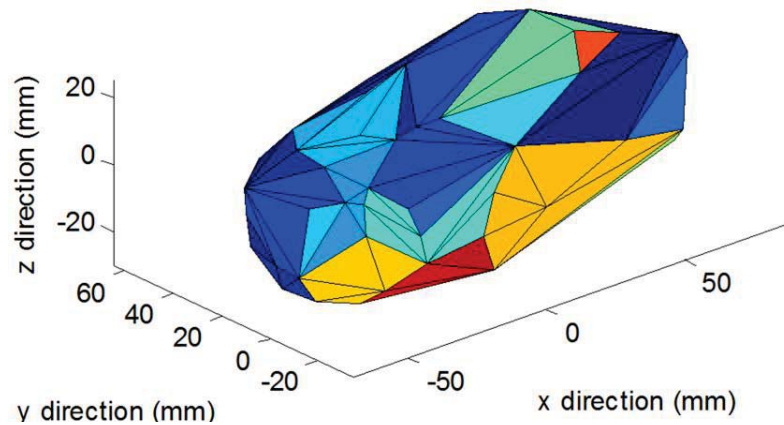


Figure 4.15. 3D model of the reconstructed mango.

4.6 CHAPTER SUMMARY

In this chapter, we described the data analyzing algorithms that extract volumetric dietary information from the structured light video. First, the overview of the food volume estimation is presented. Then, a semi-automatic algorithm is used to segment the food region from the collected videos. After that, active triangulation is used to calculate the depth maps of video frames at different directions, involving system calibration, laser cross identification, and depth map generation. The proposed ADMR can generate depth maps from the acquired data, with small errors and a large range of depths. Segmentation and depth generation results are fused together to reconstruct 3D models, resulting in a valid volume measurement. The experiments show the results for both image/video segmentation and 3D reconstruction, which prove that our algorithms are effective for data analysis.

Chapter 5. CONCLUSION AND FUTURE WORK

Chapter 5 concludes the work in this thesis and reviews the contributions. It then proposes possible directions for future work.

5.1 CONCLUSION

Dietary intake information is becoming increasingly important due to soaring obesity rates, but the current approaches to record dietary intake have severe limitations. This thesis proposed a Dietary Data Recorder System (DDRS), which integrates sensors and software to objectively measure dietary intake for free-living persons. A client-server configuration is designed for the system architecture to meet the scientific and engineering challenges of the data collection and analysis. The client consists of an ultraportable sensor package and corresponding software for data collection. The server estimates the food volume through image processing algorithms and runs the database software to manage the acquired food records. Particularly, a Mobile Structured Light System (MSLS) acquires the active 3D vision data for the food measurement. The MSLS integrates a smartphone and a laser attachment, and it has been optimized for acquiring 3D data within a small sensor package. After presenting the entire DDRS design, we have emphasized data analysis that estimates the food volume. We apply algorithms to extract food regions in the video, identify the laser pattern in the video frames, and generate depth maps using the Adaptive Distance Measurement by Regression (ADMR).

With the validation study, the proposed system has demonstrated its potential for improving the dietary intake collection. Two types of experiments have been done to test the validation of the proposed system: a system usability test and an estimation accuracy test. The system usability is tested by sending the device to clinical participants, letting them acquire dietary intake information with the device for 3 days and collecting their feedback. Ten participants have used the sensor packages for a 3-day pilot study, which was run by the Fred Hutchinson Cancer Research Center (FHCRC). Their feedback proves that the proposed system is user-friendly and improves the data recording procedure. The estimation accuracy has been tested with both the simulation and real data. The proposed algorithm could estimate the depth map

from an image with simple cross laser patterns, leading to an accurate estimated food volume. The usability and accuracy tests have proven the feasibility of the proposed approach.

5.2 FUTURE WORK

While the proposed system collects more complete data for nutritional studies and generates more objective reports compared to traditional methods, it can be further improved by future work. Possible future directions include improvements on the system integration, the sensor package, and the algorithm. For the system design, various components can be better integrated into the DDRS, particularly on the database and interface on the server side. For the sensor package, the hardware design can be improved for better data quality and portability. The laser beam can be more focused, resulting in better image quality. Besides image quality, portability can also be improved. Current housing is for concept validation, so the laser and battery are of large sizes. Smaller components can be used in the future to decrease the package size. For data analysis, better automation can be applied for the processing algorithms. Currently, manual editing is involved in the fusion of depth maps, so automating this processing can make the entire processing pipeline more user-friendly and generate more accurate results. With all of the possible research directions for future work, the proposed system will provide a better foundation for nutritional studies.

BIBLIOGRAPHY

- [1] J. Shang, et al., "A Mobile Structured Light System for Food Volume Estimation," International Workshop on Mobile Vision and International Conference on Computer Vision 2011, 2011 (accepted).
- [2] J. Shang, et al., "Dietary Intake Assessment Using Integrated Sensors and Software," IS&T/SPIE Electronic Imaging 2012, 2012 (accepted)
- [3] J. Shang, et al., "A pervasive Dietary Data Recording System," in IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011, pp. 307-309.
- [4] K. M. Flegal, et al., "Prevalence and Trends in Obesity Among US Adults, 1999-2008," JAMA: The Journal of the American Medical Association, vol. 303, pp. 235-241, January 20, 2010 2010.
- [5] H. Moller, et al., "Obesity and cancer risk: a Danish record-linkage study," Eur J Cancer, vol. 30A, pp. 344-50, 1994.
- [6] F. B. Hu, et al., "Diet, lifestyle, and the risk of type 2 diabetes mellitus in women," N Engl J Med, vol. 345, pp. 790-7, Sep 13 2001.
- [7] H. Hubert, et al., "Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study," Circulation, vol. 67, pp. 968-977, May 1, 1983 1983.
- [8] R. T. Hurt, et al., "Obesity epidemic: overview, pathophysiology, and the intensive care unit conundrum," JPEN J Parenter Enteral Nutr, vol. 35, pp. 4S-13S, Sep 2011.
- [9] E. A. Finkelstein, et al., "Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates," Health Affairs, vol. 28, pp. w822-w831, September/October 2009 2009.
- [10] R. A. Hammond and R. Levine, "The economic impact of obesity in the United States," Diabetes Metab Syndr Obes, vol. 3, pp. 285-95, 2010.
- [11] L. Kohlmeier, "Gaps in dietary assessment methodology: meal- vs list-based methods," American Journal of Clinical Nutrition, vol. 59, pp. 175S-179S, 1994.
- [12] A. H. Goris, et al., "Undereating and underrecording of habitual food intake in obese men: selective underreporting of fat intake," Am J Clin Nutr, vol. 71, pp. 130-4, Jan 2000.
- [13] L. G. Bandini, et al., "Validity of reported energy intake in obese and nonobese adolescents," Am J Clin Nutr, vol. 52, pp. 421-5, Sep 1990.
- [14] D. A. Schoeller, et al., "Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method," Can J Physiol Pharmacol, vol. 68, pp. 941-9, Jul 1990.
- [15] A. Schatzkin, et al., "A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) Study," International Journal of Epidemiology, vol. 32, pp. 1054-62, 2003.

- [16] P. Shetty, "Human energy requirements: where are we now? Issues emerging from the 2001 expert consultation on energy in human nutrition," *Food, Nutrition, and Agriculture* vol. 30, pp. 5-13, 2002.
- [17] N. Lifson and R. McClintock, "Theory of use of the turnover rates of body water for measuring energy and material balance," *Journal of Theoretical Biology* vol. 12, pp. 46-74, 1966.
- [18] W. Willett, *Nutritional Epidemiology*, 2nd ed. New York: Oxford Universal Press, 1998.
- [19] A. R. Kristal, et al., "Is it time to abandon the food frequency questionnaire?," *Cancer Epidemiology Biomarkers & Prevention*, vol. 14, pp. 2826-8, Dec 2005.
- [20] W. C. Willett and F. B. Hu, "Not the time to abandon the food frequency questionnaire: Point," *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, pp. 1757-8, 2006.
- [21] A. R. Kristal and J. D. Potter, "Not the time to abandon the food frequency questionnaire: Counterpoint," *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, pp. 1759-60, 2006.
- [22] L. Kohlmeier, "Gaps in dietary assessment methodology: meal - vs list-based methods," *American Journal of Clinical Nutrition*, vol. 59, pp. 175S-9S, 1994.
- [23] J. Garrow, "Human nutritionists' guilty secret," *BNF NUTRITION BULLETIN*, pp. 103-103, 1995.
- [24] S. Rebro, et al., "The effect of keeping food records on eating patterns," *Journal of the American Dietetic Association*, vol. 98, pp. 1163-5, 1998.
- [25] E. Barrett-Connor, "Nutrition epidemiology: how do we know what they ate?," *American Journal of Clinical Nutrition*, vol. 54, pp. 182-187, 1991.
- [26] D. Lansky, "Estimates of food quantity and calories: errors in self-report among obese patients," ed, 1982.
- [27] S. Godwin, et al., "Accuracy of reporting dietary intake using various portion-size aids in-person and via telephone," *Journal of the American Dietetic Association*, vol. 104, pp. 585-94, 2004.
- [28] L. M. Klesges, et al., "Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8-to 10-year old African-American girls: Results from the Girls Health Enrichment Multisite Studies (GEMS)," *Preventive Medicine*, vol. 38, pp. S78-87, 2004.
- [29] L. Lissner, et al., "Measuring intake in free-living human subjects: a question of bias," *Proceedings of the Nutrition Society*, vol. 57, pp. 333-9, 1998.
- [30] B. K. Armstrong, et al., *Principles of exposure measurement in epidemiology*. Oxford: Oxford University Press, 1992.
- [31] M. A. Carney, et al., "Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time " electronic interviews", " *Journal of Studies on Alcohol*, vol. 59, 1998.
- [32] M. W. Perrine, et al., "Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology," *Journal of Studies on Alcohol*, vol. 56, pp. 487-90, 1995.
- [33] R. N. Jamison, et al., "Electronic diaries for monitoring chronic pain: 1-year validation study," *Pain*, vol. 91, pp. 277-85, 2001.
- [34] S. Shiffman and A. A. Stone, "Introduction to the special section: Ecological momentary assessment in health psychology," *Health Psychology*, vol. 17, pp. 3-5, 1998.

- [35] J. Berg, et al., "Compliance with inhaled medications: the relationship between diary and electronic monitor," *Annals of Behavioral Medicine*, vol. 20, pp. 36-8, 1998.
- [36] P. W. Scruggs, et al., "Tri-Axial Accelerometry and Heart Rate Telemetry: Relation and Agreement With Behavioral Observation in Elementary Physical Education," *MEASUREMENT IN PHYSICAL EDUCATION AND EXERCISE SCIENCE*, vol. 9, p. 203, 2005.
- [37] J. E. Schwartz and A. A. Stone, "Strategies for analyzing ecological momentary assessment data," *Health Psychology*, vol. 17, pp. 6-16, 1998.
- [38] S. A. Bingham, et al., "Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records," *British Journal of Nutrition*, vol. 72, pp. 619-43, 1994.
- [39] L. Harbottle and M. B. Duggan, "Use of the PETRA scale in a weighed inventory study of British Asian children: methodological considerations," *Journal of human nutrition and dietetics(Print)*, vol. 7, pp. 275-281, 1994.
- [40] L. Stockley, et al., "Energy, protein and fat intake estimated using a food recording electronic device compared with a weighed diary," *Human Nutrition - Applied Nutrition*, vol. 40, pp. 19-23, 1986.
- [41] M. J. Kretsch and A. K. Fong, "Validation of a new computerized technique for quantitating individual dietary intake: the Nutrition Evaluation Scale System (NESSy) vs the weighed food record," *American Journal of Clinical Nutrition*, vol. 51, pp. 477-84, 1990.
- [42] C. H. Lindquist, et al., "Use of Tape-Recorded Food Records in Assessing Children's Dietary Intake," *Obesity Research*, vol. 8, pp. 2-11, 2000.
- [43] D. A. Williamson, et al., "Comparison of digital photography to weighed and visual estimation of portion size," *Journal of the American Dietetic Association*, vol. 103, pp. 1139-45, 2003.
- [44] C. Sun, et al., "Thickness measurement and crease detection of wheat grains using stereo vision," *Pattern Recogn. Lett.*, vol. 28, pp. 1501-1508, 2007.
- [45] K. Imou, et al., "Three-dimensional shape measurement of strawberries by volume intersection method," *Transactions of the American Society of Agricultural Engineers*, vol. 49, pp. 449-456, 2006.
- [46] D. J. Lee, et al., "Fast surface approximation for volume and surface area measurements using distance transform," *Optical engineering*, vol. 42, pp. 2947-2955, 2003.
- [47] G. P. Moreda, et al., "Non-destructive technologies for fruit and vegetable size determination - A review," *Journal of Food Engineering*, vol. 92, pp. 119-136, 2009.
- [48] L. E. Burke, et al., "Self-monitoring dietary intake: current and future practices," *Journal of Renal Nutrition*, vol. 15, pp. 281-90, 2005.
- [49] J. Beasley, et al., "Accuracy of a PDA-based dietary assessment program," *Nutrition*, vol. 21, pp. 672-7, 2005.
- [50] B. A. Yon, et al., "The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake," *Journal of the American Dietetic Association*, vol. 106, pp. 1256-9, 2006.
- [51] C. H. Kaczowski, et al., "Four-Day Multimedia Diet Records Underestimate Energy Needs in Middle-Aged and Elderly Women as Determined by Doubly-Labeled Water " *Journal of Nutrition*, vol. 130, pp. 802-5, 2000.

- [52] D.-H. Wang, et al., "Development of a new instrument for evaluating individuals' dietary intakes," *Journal of the American Dietetic Association*, vol. 106, pp. 1588-93, 2006.
- [53] C. K. Martin, et al., "Quantification of food intake using food image analysis," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 6869-6872.
- [54] Z. Fengqing, et al., "The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, pp. 756-766, 2010.
- [55] R. Weiss, et al., "Automatic Food Documentation and Volume Computation Using Digital Imaging and Electronic Transmission," *Journal of the American Dietetic Association*, vol. 110, pp. 42-44, 2010.
- [56] M. Sun, et al., "A wearable electronic system for objective dietary assessment," *Journal of the American Dietetic Association*, vol. 110, pp. 45-7, 2010.
- [57] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, pp. 150-162, 1994.
- [58] S. K. Nayar and Y. Nakagawa, "Shape from focus," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, pp. 824-831, 1994.
- [59] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," in *Shape from shading*, K. P. H. Berthold, Ed., ed: MIT Press, 1989, pp. 513-531.
- [60] R. Szeliski, "Computer Vision: Algorithms and Applications," 2010.
- [61] http://en.wikipedia.org/wiki/Pinhole_camera_model.
- [62] J. Salvi, et al., "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, pp. 827-849, 2004.
- [63] M. Levoy, et al., "The digital Michelangelo project: 3D scanning of large statues," presented at the ACM SIGGRAPH, 2000.
- [64] A. Banno, et al., "Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project," *International Journal of Computer Vision*, vol. 78, pp. 207-222, 2008.
- [65] B. Trinchão Andrade, et al., "3D preserving xviii century barroque masterpiece: Challenges and results on the digital preservation of Aleijadinho's sculpture of the Prophet Joel," *Journal of Cultural Heritage*, vol. In Press, Corrected Proof.
- [66] Y. Bok, et al., "Capturing Village-level Heritages with a Hand-held Camera-Laser Fusion Sensor," *International Journal of Computer Vision*, vol. 94, pp. 36-53, 2011.
- [67] T. van Lankveld, et al., "Identifying rectangles in laser range data for urban scene reconstruction," *Computers & Graphics*, vol. 35, pp. 719-725, 2011.
- [68] L. Zhang, et al., "Spacetime faces: high resolution capture for modeling and animation," presented at the ACM SIGGRAPH, Los Angeles, California, 2004.
- [69] H. Kawasaki, et al., "Dense 3D Reconstruction with an Uncalibrated Active Stereo System," in *Asian Conference on Computer Vision*, ed, 2006, pp. 882-891.
- [70] H. Hiroshi Kawasaki, et al., "Dynamic scene shape reconstruction using a single structured light pattern," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [71] R. Szeliski, *Computer Vision: Algorithms and Applications*: Springer, 2010.

- [72] F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, vol. 13, pp. 231-243, 2004.
- [73] Z. M. Bi and L. Wang, "Advances in 3D data acquisition and processing for industrial applications," *Robotics and Computer-Integrated Manufacturing*, vol. 26, pp. 403-413, 2010.
- [74] J. Salvi, et al., "A review of recent range image registration methods with accuracy evaluation," *Image Vision Computing*, vol. 25, pp. 578-596, 2007.
- [75] S. K. Mada, et al., "An overview of passive and active vision techniques for hand-held 3D data acquisition," *Proceedings of SPIE*, vol. 4877, 2003.
- [76] D. Lanman and G. Taubin, "Build your own 3D scanner: 3D photography for beginners," presented at the ACM SIGGRAPH 2009 Courses, New Orleans, Louisiana, 2009.
- [77] D. Lanman, et al., "Surround Structured Lighting for Full Object Scanning," presented at the Proceedings of the Sixth International Conference on 3-D Digital Imaging and Modeling, 2007.
- [78] S. F. Schakel, *Maintaining a nutrient database in a changing marketplace: Keeping pace with changing food products-A research perspective*. Kidlington, ROYAUME-UNI: Elsevier, 2001.
- [79] L. Xu and Y. Li, "Optimization design for structured light system," in *IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS)*, 2010, pp. 602-605.
- [80] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 603-619, 2002.
- [81] J. Ning, et al., "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognition*, vol. 43, pp. 445-456, 2010.
- [82] L. Guisser, et al., "PGSD: an accurate 3D vision system using a projected grid for surface descriptions," *Image and Vision Computing*, vol. 18, pp. 463-491, 2000.
- [83] F. S. Marzani, et al., "Calibration of a three-dimensional reconstruction system using a structured light source," *Optical Engineering*, vol. 41, pp. 9484-492, 2002.
- [84] A. Dipanda and S. Woo, "Towards a real-time 3D shape reconstruction using a structured light system," *Pattern Recognition*, vol. 38, pp. 1632-1650, 2005.

VITA

Junqing Shang was born in Taiyuan, Shanxi, China. He is currently studying in Electrical Engineering at the University of Washington, Seattle, WA. His research interests are in the areas of signal and image processing, computer vision, and machine learning.