A Novel 3D Sensory System for Robotic Urban Search and Rescue Missions

by

Babak Mobedi

A thesis submitted in conformity with the requirements for the degree of Masters of Applied Science
Mechanical and Industrial Engineering
University of Toronto

© Copyright by Babak Mobedi 2010
A Novel 3D Sensory System for Robotic Urban Search and Rescue Missions

Babak Mobedi
Masters of Applied Science
Mechanical and Industrial Engineering
University of Toronto
2010

Abstract

Urban Search and Rescue (USAR) is the emergency response function that deals with the collapse of man-made structures. USAR environments contain concrete rubble, dust and debris, and provide poor lighting conditions. Due to the dangers that USAR rescue workers and their canines face, robots have become of interest in aiding rescue workers in searching. Experiences with robots in USAR missions have shown that a compact 3D sensor for 3D mapping of the environment is beneficial in providing the robot and identified victims’ locations within the structurally unstable environment. This thesis presents the developments of a novel 3D sensory system that provides both 3D and 2D texture information for mapping of cluttered unknown USAR environments. The sensor has been integrated into a robot platform, and experiments conducted to validate its usability in such applications. The experimental results show the potential for using this sensor in USAR robot mission.
Acknowledgments

I would like to sincerely thank my advisor, Professor Goldie Nejat, for her kind, yet constructive guidance and support of my research work. I gratefully thank Professor Beno Benhabib for kindly providing me with access to his laboratory to use the linear-stage utilized for calibration and measurement error estimation of my 3D sensory system. I would like to thank Professor Douglas Hooton for providing concrete rubble to build the USAR-like scene. I would like to thank my M. A. Sc. thesis committee for their time and input. I would like to thank Onome Igharoro for the development of a GUI for the mobile robot platform, Bijan Shahriari for his help with system calibration, and Lawrence Lee for the development of the 3D sensory system enclosure. In addition, I would like to thank Hemant Kumar for the design of a vibration isolation system for mounting the 3D sensory system, Jayiesh Singh for building the robot power system circuitry, and Bill Liu for his assistance with the 3D mapping algorithm. Lastly, I would like to thank all of my lab mates for their motivation and support during the two year period of my research.
# Table of Contents

Acknowledgments.......................................................................................................................... iii

Table of Contents........................................................................................................................... iv

List of Figures ................................................................................................................................... vii

Chapter 1 Introduction .................................................................................................................... 1

1.1 Motivation.................................................................................................................................. 1

1.2 Literature Review...................................................................................................................... 3

    1.2.1 Stereo Vision Sensors ............................................................................................. 3

    1.2.2 3D Laser Range Finders .................................................................................... 5

    1.2.3 3D Time-of-Flight Cameras .................................................................................. 6

1.3 Problem Definition.................................................................................................................... 7

1.4 Proposed Methodology and Tasks.............................................................................................. 8

    1.4.1 Structured Light Sensing ....................................................................................... 8

    1.4.2 3D Sensory System ............................................................................................... 8

    1.4.3 Implementation ....................................................................................................... 9

    1.4.4 Conclusion .............................................................................................................. 9

Chapter 2 Structured Light Sensing.............................................................................................. 10

2.1 Motivation.................................................................................................................................. 10

2.2 Structured Light Sensory Systems............................................................................................. 11

    2.2.1 Single Coded Pattern Techniques ........................................................................... 11

    2.2.2 Multiple Coded Pattern Techniques ...................................................................... 15

2.3 Phase-Shifting Techniques........................................................................................................ 16

    2.3.1 Sinusoidal Phase-Shifting Technique .................................................................... 18

    2.3.2 Triangular Phase-Shifting Technique .................................................................... 21

    2.3.3 Trapezoidal Phase-Shifting Technique ................................................................. 24
2.4 Phase Unwrapping ............................................................................................................ 27
  2.4.1 Path Dependent Techniques .................................................................................. 27
  2.4.2 Path Independent Techniques ............................................................................... 31
2.5 Chapter Summary ............................................................................................................. 34

Chapter 3 3D Sensory System ............................................................................................ 35
  3.1 Hardware Development ............................................................................................... 35
  3.2 Software Development ............................................................................................... 37
    3.2.1 Active Phase Unwrapping ..................................................................................... 39
    3.2.2 Shadow Detection ................................................................................................. 43
    3.2.3 Capture Modes ...................................................................................................... 45
  3.3 System Calibration ....................................................................................................... 46
    3.3.1 Colour Calibration ................................................................................................ 46
    3.3.2 Intrinsic Parameter Estimation .............................................................................. 54
    3.3.3 Extrinsic Parameter Estimation ............................................................................. 55
  3.4 Chapter Summary ....................................................................................................... 58

Chapter 4 Experiments ........................................................................................................... 59
  4.1 Measurement Error Estimation .................................................................................... 59
    4.1.1 Depth (z-axis) Error .............................................................................................. 59
    4.1.2 X- and Y- axis Errors .......................................................................................... 62
  4.2 Object Surface Measurement ...................................................................................... 64
  4.3 Object Surface Effects on Measurements .................................................................... 66
  4.4 Shadow Detection ...................................................................................................... 67
  4.5 Illumination Conditions .............................................................................................. 68
  4.6 3D Mapping ................................................................................................................. 70
    4.6.1 3D Map Building Results ..................................................................................... 73
    4.6.2 Recall Rate ............................................................................................................ 73
4.7 Discussion of Experimental Results ................................................................. 78
4.8 Chapter Summary ............................................................................................. 81
Chapter 5 Conclusion ............................................................................................ 83
  5.1 Summary of Contributions ............................................................................. 83
    5.1.1 A 3D Structured Light Sensor for USAR Missions .............................. 83
    5.1.2 Implementation ..................................................................................... 84
  5.2 Discussion of Future Work ........................................................................... 85
  5.3 Final Concluding Statement ....................................................................... 86
References ............................................................................................................ 87
Appendices ........................................................................................................ 92
  A.1 Intrinsic Parameter Estimation Results ....................................................... 92
  A.2 Extrinsic Parameter Estimation Results ..................................................... 93
List of Figures

Figure 1: Stereo-vision system....................................................................................................... 4
Figure 2: Structured Light Hardware........................................................................................... 10
Figure 3: A Moiré Fringe transducer [49].................................................................................... 12
Figure 4: Moiré Fringe Pattern using two straight-line gratings.................................................. 12
Figure 5: Deformed moiré fringes on a model air-plane [33]...................................................... 13
Figure 6: Colour-encoded fringe pattern: (a) Projected pattern, (b)-(d) Captured images of the three sinusoidal phase-shifted patterns (phase shifted by $2\pi/3$ rad), and (e) 3D surface image [35]............................................................................................................................................... 14
Figure 7: Binary-coded patterns [36]........................................................................................... 15
Figure 8: Rubble scene................................................................................................................. 17
Figure 9: Directly coded projector phase image........................................................................... 18
Figure 10: Sinusoidal phase-shifted fringe patterns: (a) Pattern 1, (b) Pattern 2, and (c) Pattern 3. ....................................................................................................................................................... 19
Figure 11: Modulo $2\pi$ image using sinusoidal phase-shifted fringe patterns ......................... 20
Figure 12: Triangular phase-shifted fringe patterns: (a) Pattern 1, and (b) Pattern 2................. 23
Figure 13: Defocused fringe of two-step triangular phase-shifting method [29]....................... 24
Figure 14: Trapezoidal phase-shifted fringe patterns: (a) Pattern 1, (b) Pattern 2, and (c) Pattern 3.................................................................................................................................................... 26
Figure 15: Defocused fringe of two-step triangular phase-shifting method [28]....................... 27
Figure 16: Pixel queuing: (a) Comparison of pixels, and (b) Unwrapping path [52]............... 28
Figure 17: Tiled section of phase unwrapped image [53]............................................................ 29
Figure 18: Quality-guided map algorithm’s path [54]................................................................. 30

Figure 19: Quad-tree decomposition method [58]................................................................. 32

Figure 20: Region-referenced window patching method [59]..................................................... 33

Figure 21: 3D Sensory System Architecture. ............................................................................. 35

Figure 22: 3D Sensory System Hardware .................................................................................. 36

Figure 23: Projector and Trigger board signal ........................................................................ 37

Figure 24: Captured images of projected patterns on the rubble scene: (a)-(c) Set 1, and (d)-(f) Set 2. ............................................................................................................................................. 38

Figure 25: Modulo $2\pi$ image of Set 2. .................................................................................. 40

Figure 26: Phase unwrapped image: (a-c) active phase unwrapping technique, and (d-f) quality-guided map technique. .................................................................................................................. 42

Figure 27: Textured 3D information obtained using: (a-c) the active phase unwrapping technique, and (d-f) the quality-guided map technique. ................................................................................. 43

Figure 28: Shadow and occluded regions. ................................................................................ 43

Figure 29: Acquisition mode range diagram. ........................................................................... 46

Figure 30: Intensity response evaluation: (a) pre-calibration, and (b) post-calibration.......... 48

Figure 31: Linear intensity response error: (a) Red channel, and (b) Green channel .......... 49

Figure 32(a): Short-range mode intensity response evaluation for pre-calibration. ............... 51

Figure 33: Linear intensity response error ............................................................................... 52

Figure 34: Short-range mode intensity response evaluation: (a) pre-calibration, and (b) post-calibration. ............................................................................................................................................. 53

Figure 35: Linear intensity response error ............................................................................... 54
Figure 36: error caused by nonlinear image distortion for: (a) the camera, and (b) the projector.

Figure 37: System Coordinate Frames.

Figure 38: Common coordinate frame: (a) Camera image, and (b) Projector image.

Figure 39: Experimental setup: (a) overall set-up, (b) flat wooden board used in testing, and (c) checkerboard pattern used in testing.

Figure 40: USAR-like scene.

Figure 41: Distance measurement error within the 3D sensor’s working range.

Figure 42: 3D point cloud of flat object at 350mm range. (Note the effects of saturated camera pixels).

Figure 43: Measurement error throughout sensor working range: (a) x-axis error, and (b) y-axis error.

Figure 44: Surface Measurement Objects: (a) 3” Gauge block, and (b) Cylindrical weight.

Figure 45: Surface profile cross-section measurements of: (a) A flat surface (gauge block), and (b) Curved surface (cylindrical weight).

Figure 46: 2D and 3D information of objects with various material compositions and surface properties: (a) Plastic fan, (b) Concrete, (c) Wooden board, (d) Aluminum plate, (e) Human face, and (f) Mirror. 3D point clouds representing objects close to the sensor appear white, whereas objects further away appear gray.

Figure 47: 2D texture, 2D texture with masked shadow pixels (shown in cyan), and 3D textured point cloud information of objects with 3D sensor located at various poses.

Figure 48: 2D texture, 2D texture with masked shadow pixels, and 3D textured point cloud information of objects with various ambient lighting conditions: (a) dark condition, (b) ambient fluorescent room lighting, and (c) bright lighting condition using strobe light.

Figure 49: Identified SIFT Keypoints in the first pair of 2D texture images.
Figure 50: Matched SIFT Keypoints in the first pair of 2D texture images. ................................. 71

Figure 51: Stitching of the first three 3D point clouds. ................................................................. 72

Figure 52(a): 3D map with superimposed robot path. ................................................................. 74

Figure 53: Zoom-in views 3D map. ............................................................................................... 76

Figure 54: Scale Invariant Feature Transform (SIFT) recall rate evaluation. ............................ 77

Figure 55: Pose change influence on SIFT keypoint descriptors: (a) Identified SIFT keypoint, (b) Rotation of optical axis, and (c) Translation of optical axis. ................................................. 78

Figure 56: CCD and DMD chip geometric representation: (a) Less influential alignment, and (b) More influential alignment. ........................................................................................................ 80

Figure 57: Illumination reduction illustration. Object distance is: (a) close to the sensor, and (b) far from the sensor. .............................................................................................................. 80
Chapter 1
Introduction

1.1 Motivation

The recent earthquakes that hit Haiti and Chile in early 2010 are another harsh reminder of the impact of natural disasters on infrastructure, and more importantly human lives. The Centre for Research on the Epidemiology of Disasters (CRED) defines a disaster as a situation or event, which overwhelms local capacity, necessitating a request to national or international level for external assistance; an unforeseen and often sudden event that causes great damage, destruction and human suffering [1]. From 1999-2008, the total cost of damage caused by world disasters has been estimated to be over $1 trillion US dollars [2]. Furthermore, these disasters have killed over 1.2 million people and affected more than 2.6 million others. In urban disasters, a trapped victim’s survival rate can drastically drop from 91% within the first 30 minutes to 33.7% after 2 days [3]. Therefore, timely response of rescuing these victims is crucial in increasing their survival chances.

Urban Search and Rescue (USAR) is defined as the emergency response function that deals with the collapse of man-made structures [4]. USAR scenes can contain piles of concrete rubble, exposed metal, dust and debris due to collapsed buildings. Poor ambient lighting conditions and the structural instability of the environment are additional obstacles that USAR rescue workers are faced with. Currently, canine dogs are commonly used for aiding rescue workers in searching for victims. However, due to the harsh environmental conditions that are common to USAR scenes, such as hot areas due to fire or small voids, the use of canines has its limitations. Furthermore, these rescue efforts can take up many days, making the fatigue of canines and human rescue workers a significant issue. As a result, robots have become of interest in aiding rescue workers for searching areas that are too hot, or narrow for rescue workers and canines and relieving them from of their tiresome duties. In 1995, the earth quake in Kobe, Japan and the bombing in Oklahoma City, US triggered the use of rescue robots to aid rescue workers in finding survivors in Urban Search and Rescue (USAR) missions [5]. The Central Disaster Prevention Committee of Japan included the development of robots and systems that work in inaccessible areas as an intended area of development [6]. In the US, personal experiences of J. Blitch, a Master’s of Science student of R. R. Murphy, as a rescue worker in the Oklahoma City
bombing lead to his and R. R. Murphy’s commitment to a new field of research that is USAR robotics [5]. Thus far, rescue robots have been used in a number of USAR efforts, such as the World Trade Center (WTC) disaster scene in 2001, the earthquake in Chuetsu, Niigata, Japan in 2004, and Hurricanes Katrina, Rita, and Wilma that hit the Southern Gulf coast of United States in 2005 [7].

In the WTC disaster environment, robots carrying cameras were utilized to find victims by entering the scene through voids too small or deep for a person to enter. The robots were also used to survey larger voids that people were not permitted to enter as a result of fire and/or structural instability [8]. The robot operators used the robots’ cameras to navigate a robot, detect objects and victims, and recover a robot from the scene [9]. Due to environmental limitations, in particular the challenging terrain and the high heat sources within the rubble piles, the robots were restricted in the tasks that they could accomplish. The use of only 2D video cameras made the task of object identification very difficult; resulting in cases where materials such as concrete and metal were mistaken for victims [9]. Furthermore, lack of 3D sensory information also hindered a robot operator’s situational awareness resulting in the robots getting stuck more than 2 times per minute [9]. Situational awareness is defined as the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future [10]. Since then, in addition to 2D cameras, robots in more recent efforts have been equipped with thermal cameras, CO2 sensors and acoustic cameras for victim identification [7]. Unmanned Air Vehicles (UAVs) and a VGTV Extreme robot were used in rescue efforts after hurricane Katrina. The VGTV Extreme robot was used to search the first floor of an apartment building, which was unsafe for human or canine entry due to gas leaks and the unstable nature of the structure [11]. The VGTV Extreme was equipped with an NTSC camera and two halogen lights. A successful search revealed that there were no victims in the apartment. Post disaster experimentation was conducted after the Cheutsu earthquake. The International Rescue System Institute (IRS) utilized the Soryu III robot developed in collaboration with the Tokyo Institute of Technology to search a house destroyed by the earthquake. The robot was equipped with a CCD (Charge-Coupled Device) camera, an infrared camera, and a two-way audio [6]. The testing in the debris concentrated on mobility.

Experiences gained through the use of robots in USAR efforts revealed that rescue robots were missing an important primary sensor: a compact 3D sensor that can potentially be used to
generate a 3D map of a disaster environment [7]. Such information is essential in providing both the robot’s location as well as the location of identified victims within the environment, thereby improving the robot operator’s situational awareness and reducing his/her cognitive load when navigating the robot and searching for victims. Furthermore, the use of 3D information of objects can reduce false object detection, as it allows the operator to view an object of interest from various perspectives; the terrain of a USAR environment may not allow the operator to easily navigate the robot around an object of interest when using a 2D camera.

1.2 Literature Review

Although 3D sensors have yet to be integrated into mobile robots for USAR missions, a number of research efforts have been focused on developing and testing such sensors in similar environments. Current 3D sensors integrated into USAR robots fall under three categories: stereo vision, laser range finders, and 3D time-of-flight (TOF) cameras. The following subsections discuss the various types of 3D sensors developed for obtaining 3D information from a USAR scene.

1.2.1 Stereo Vision Sensors

Stereo vision sensors use two or more cameras to obtain 3D information of a scene [12]. Consider the stereo vision setup shown in Figure 1. Each camera captures an image of the scene. Common points, such as point P shown as red in Figure 1, are identified and their pixel coordinates in each camera’s image, labelled P(i1, j1) and P(i2, j2), are used along with triangulation to obtain their 3D coordinate measurements. Triangulation uses the distance between the two cameras, the angle of point P’s pixel in Camera 1 with respect to its optical axis, the angle of point P’s pixel in Camera 2 with respect to its optical axis to obtain the 3D coordinate measurements of point P.
Figure 1: Stereo-vision system.

In [13, 14], a stereo vision system is developed for USAR applications using two web cameras (with a resolution of 320x240 pixels). In [13], a region-based approach for stereo matching is proposed rather than using individual pixels. Extraction and matching of regions is computationally expensive resulting in a frame rate of 1.5-2.5 fps with the current hardware. Thus far, the technique has been evaluated in a structured office environment consisting of computers, desks, and chairs and has yet to be applied to 3D scene mapping. In [14], a line-segment extraction for semi-dense stereo matching is proposed. Extracting line-segments and matching them occurred at 0.66-2.22fps. 3D measurements of scenes with various objects are presented. Compared with the presented ground truth depth image, the measurement results using their technique showed errors in various parts of the scene. Furthermore, the 3D data for line-segments is only obtained; merging the split line segments to fill the rest of the image with 3D data points is stated as part of a future development.

In general, stereovision techniques are highly dependent on cooperative surfaces, mainly on the presence of surface textures and on ambient light [13]. Thus, their performance may suffer in USAR scenes where the object surfaces may not provide sufficient texture information, or lighting.
1.2.2 3D Laser Range Finders

Laser range finders use a laser as the light source and rotate a spot or line onto the scene to be observed from a sensor. The relative positions and orientations of the laser and sensor are used along with triangulation to obtain the distance of the point(s) to the sensor [12].

In [15-17], 3D laser range finders were developed utilizing a 2D laser scanner and a rotating mount. In [15], the developed 3D laser range finder was able to provide a single scan with a resolution of 361x176 points. Multiple scans were taken by the laser range finder and merged to generate a 3D map of the 2004 RoboCup Rescue 2004 arena. In [16], a small 2D laser range finder is rotated to obtain 3D sensory information of indoor environments with smooth terrains. In contrast to the laser range finder in [15], the 3D laser scanner in [16] provides 3D measurements at 40fps, however, at a lower resolution of 1,052 3D points per scan. In [17], a 2D laser range finder is rotated 360° to obtain 3D range images of an indoor environment with a resolution of 1,080 data points per acquisition at 0.025fps. In [18, 19], three prototypes of a laser range finder consisting of an infrared laser module with a CCD camera to capture the laser beam and another camera for capturing 2D texture images, are presented. The first prototype consists of a laser module, a conical mirror, a hyperbolic mirror and a CCD camera. For the second prototype, a second CCD camera is added for capturing texture information. The third prototype is a more compact version of the second prototype, with the two CCD cameras exchanged with a progressive camera, and a CMOS camera. LEDs are also used around the laser module to brighten the ambient illumination conditions. The three prototypes were able to obtain 360 data points per acquisition. Their optimal range was determined to be 300mm. Experiments on the third prototype were conducted in two different environments. Results from the first environment showed good measurements of a path way’s geometric shape with the exception of side walls that appeared wavy because of their large distance to sensor. The piece of wood was well reconstructed; however, the plastic bottle could not be recognized in the 3D reconstructed image due to measurement errors cause by the specularity of its surface. The results of the USAR-like scene showed good quality measurements of the doll’s face and hand.

The main limitation of these types of sensor is that they do not directly provide texture information of the scene and therefore rely on robot internal sensors (encoders, accelerometers, and gyro sensors) or time consuming algorithms that only use 3D coordinate measurements for
mapping. Laser range finder systems that do provide texture information, such as the one described in [18, 19], utilize a separate 2D camera to obtain texture information. This requires matching 3D and 2D information, which due to the different resolution of the two components, is not a one-to-one correspondence and requires further calibration. Furthermore, the resolution of the component with the lowest resolution limits the resolution of the 2D texture image. Lack of texture information also hinders object recognition and victim identification. In addition, another limitation of laser range finders is specular reflections. When the light reflects from polished metal or glass, the reflected beam behaves in an unpredictable manner, causing incorrect range measurements [12].

1.2.3 3D Time-of-Flight Cameras

3D TOF cameras use time-of-flight technology to obtain 3D surface information of objects in a scene. They illuminate the scene using modulating infrared LEDs [20]. The reflected light is then captured by a CCD/CMOS sensor. By measuring the time that the modulated light pulses take to travel to the objects and back onto the sensor, the range information of objects in the scene can be obtained.

In [21-23], the CSEM Swiss Ranger 3D camera with a resolution of 160x124 pixels was used for 3D mapping of USAR-like scenes. In [21], the sensor was used along with a 2-axis accelerometer (to measure pitch and roll) to develop a 3D mapping technique that was tested in a simulated collapsed house environment. It was found that the sensor was not able to provide short range information and that a large error was present in the yaw rotation of the robot. Furthermore, the presence of sporadic noise and its increase as the distance from the camera to the scene increased, limited the working range of the 3D sensor to 500-3500mm. The low resolution of the sensor also made it difficult to identify the majority of the objects in the scene.

In [22], in addition to the CSEM Swiss Ranger 3D camera, a 2D camera and a thermal camera were used for obtaining texture and heat information. Due to the resolution limitations, the 3D camera was chosen as the reference sensor for the data to be merged with. As a result, the resolution of the 2D texture and thermal images was reduced to the 160x124 pixel resolution of the 3D sensor. In [23], a 2D camera with a 640x480 resolution is used along with the CSEM Swiss Ranger 3D camera. It was stated that 3D coordinate measurements obtained from specular surfaces was noisy. In [24], a newer model of the Swiss Ranger camera (SR-3000 with a pixel
resolution of 176x144) was used along with a 2D camera (with a resolution of 1024x768 pixels) for mapping USAR environments. The 3D coordinate measurements of objects with specular surfaces were still noisy. In addition, it was stated that ambiguous measurements are obtained when measuring beyond the sensor’s maximum range of 7500mm. In [25], the Canesta Range Sensor with a resolution of 64x64 pixels was evaluated for USAR mobile robot navigation and mapping applications. Experiments were conducted in both indoor and outdoor environments. It was shown that objects, i.e. black plastic and water, that absorb the emitted wavelength appear invisible to the sensor. The sensor also had a limitation in identifying objects that are outside of its working range. Due to the ambiguity of distance measurements beyond the working range of the sensor, objects that are, i.e. 200mm, beyond the maximum range appear 200mm from the minimum range of the sensor. In addition, the low resolution of the 3D information can hinder the accuracy of stitching 3D images to build a 3D map.

In general, the pixel array size of current 3D TOF cameras is limited and hence the resolution of the 3D information can be low. They can also experience focal blurring due to the interference of infrared rays reflected back from objects as the distance to the objects increases [20]. Furthermore, as the object distance to the sensor increases, aliasing of object measurements may also occur [22].

The aforementioned 3D sensors are an improvement over 2D sensors for obtaining detailed sensory information of a USAR scene using a mobile robot. However, the limitations that are inherent to each type of these sensors, such as lack of 2D texture information corresponding to 3D coordinate measures, inability to obtain measurements in dark environments, and low resolution, necessitate the development of a new 3D sensor for USAR robot applications. In particular, the design of the 3D sensor should aim to provide 3D and 2D texture information of the scene with sufficient resolution for navigation, and object and victim identification.

### 1.3 Problem Definition

This thesis focuses on the design of a new 3D sensor for robotic mapping and localization in cluttered USAR environments. These types of environments tend to be dark or dimly lit consisting of various objects with different surface types and colours. Furthermore, these objects can be covered with dust, providing little texture information. Such environments present challenges for current 3D sensors. Stereo vision based systems depend on the presence of texture
information of objects and sufficient ambient illumination to obtain 3D coordinate measurements. 3D Laser finders and 3D TOF cameras do not suffer from this setback; however, they cannot directly provide 2D texture information of 3D coordinate measurements. This is sometimes done using an additional camera. Furthermore, 3D TOF cameras lack adequate resolution and cannot identify measurements beyond their maximum range.

The design focus of this thesis is to develop a new 3D sensor to address these limitations. The 3D sensor is to provide 3D coordinate measurements with corresponding 2D texture information. Furthermore, it is to provide 3D images with sufficient resolution of dark, as well as bright indoor environments.

1.4 Proposed Methodology and Tasks

The overall design of the 3D sensor comprises the following components with corresponding reference to the Thesis Chapters:

1.4.1 Structured Light Sensing

In Chapter 2, a discussion on structured light techniques is presented as motivation for the use of this type of technique for the proposed 3D sensor. The active nature of structured light systems and their ability to provide 3D coordinate measurements with corresponding 2D texture information makes them suitable for USAR applications. A brief overview of current structured light techniques including their abilities and limitations is provided. The chapter focuses on the phase-shifting technique as the proposed technique for structured light sensing in USAR environments. An overview on phase-unwrapping, a significant step in obtaining 3D information using the phase-shifting technique is presented, as well as discussion on its limitation in USAR environments.

1.4.2 3D Sensory System

In Chapter 3, the development of the proposed 3D sensory system is presented including both software and hardware components. A calibration procedure is proposed for the overall sensory system. With regards to system software development, a new Active Phase Unwrapping technique is proposed to address the limitations of previous phase unwrapping techniques in
USAR environments. In addition, an occlusion detection algorithm is proposed to identify and eliminate 3D information corresponding to regions of shadows within the scene.

1.4.3 Implementation

In Chapter 4, extensive experiments are presented to evaluate the performance of the developed 3D sensor in various conditions for USAR robot missions. Discussions and comparisons to illustrate the effectiveness of the proposed approaches are also presented.

1.4.4 Conclusion

Finally, Chapter 5 presents concluding remarks on the development of the 3D sensor, highlighting the main contributions of the thesis and future work.
Chapter 2
Structured Light Sensing

2.1 Motivation

In general, structured light systems consist of one (or more) camera(s) and active light source(s), where the latter project a known light pattern onto an object of interest that the camera(s) can capture for 3D reconstruction [26], Figure 2. In particular, the captured image(s) of projected pattern(s) are matched to obtain correspondence between the camera and projector [27]. Triangulation is then used to determine the 3D information of each pixel in the image.

![Structured Light Hardware](image)

**Figure 2: Structured Light Hardware.**

The active nature of structured light sensing makes this type of sensor suitable for dark/dimly lit environments that are common to USAR operations. Determining the correspondence information of projected patterns from captured images is relatively simpler than determining correspondence information using stereo vision based systems, which match pixels between two captured images [28]; lack of surface texture information due to dust further complicates pixel matching for stereo vision based systems. When compared with laser range finders, structured light sensors are capable of obtaining 3D coordinate measurements of an area based on their field-of-view, rather than a single point or an array of points along a laser line. This allows structured light sensors to obtain a large amount of measurements at a faster rate than laser range finders. 3D TOF cameras in comparison to structured light sensors typically lack in resolution and the ability to provide 2D texture information of the 3D coordinate measurements obtained.
Current structured light sensors are commonly used in applications in which 3D surface information of a single stand alone object is required such as human faces [28], historical artefacts [29, 30], and surface inspection of manufactured parts [31]. They are yet to be applied to applications where the 3D coordinate measurements of scenes consisting of multiple objects, oriented in an unknown and unstructured manner, are required. Therefore, the development of the 3D sensor based on structured light requires the adaptation of this technique for scenes such as USAR environments where multiple objects exist.

2.2 Structured Light Sensory Systems

Structured light approaches can be broadly categorized into two types of techniques: single and multiple coded patterns. Single coded patterns project only one pattern from the light source to be captured by the camera, whereas multiple coded patterns project a series of patterns from their light source and quickly switch between them to capture in real-time. Examples of single coded pattern approaches include the Moiré technique [32, 33], the rainbow 3D camera method [34], and the colour-encoded fringe projection technique using a Digital Light processing (DLP) projector [35]. Multiple coded pattern approaches mainly include the binary-coded technique [36, 37] and the phase-shifting technique [28, 29, 38-47]. The following subsections discuss the details of both of these structured light approaches.

2.2.1 Single Coded Pattern Techniques

The following subsections discuss the aforementioned examples of single coded pattern techniques.

2.2.1.1 Moiré Technique

The term “Moiré” in optics refers to a beat pattern produced between two gratings of approximately equal spacing [48]. Figure 3 shows the hardware setup of a moiré fringe transducer for sensing the level of a surface. Figure 4 shows a moiré pattern produced by two straight-line gratings that are rotated by a small angle with respect to each other. As the angle between the two gratings increases, the number of fringes that appear in the pattern increases as well. When the projected pattern is viewed from an angle, the moiré fringes appear deformed, as shown in Figure 5. The deformed shape of the fringes is relative to the surface contour of the
object and the view angle with respect to the light source. The deformation information of the fringes can be used to obtain the 3D surface contour of the object.

Figure 3: A Moiré Fringe transducer [49].

Figure 4: Moiré Fringe Pattern using two straight-line gratings.
In [33], the generation of 3D surface contours using the moiré fringe patterns is presented. A camera is placed vertically above a square wave grid, with two light sources equidistant from the camera. The objects to be measured were placed under the grid. A model airplane, a model car, and a conical object were contoured (measured) in the experiments. It’s stated that the depth of field of the technique is limited by diffraction, and the method produces noise due to the grid lines used.

In [32], a method for moiré fringe generation is discussed that allows for differentiation of object surface slope direction. In general, traditional moiré techniques can differentiate between the height changes on an object’s surface; however, the direction of height change is ambiguous. The proposed method does not provide absolute slope determination, however can provide the relative slope direction of one surface with respect to any other slope on the same continuous surface. The relative slope direction obtained from this technique can be used in combination with the traditional moiré techniques to determine the absolute slope of an object’s surface.

### 2.2.1.2 Rainbow 3D camera

In [34], a rainbow 3D camera is presented consisting of a camera, a white light source, a cylindrical lens and a Linear Variable Wavelength Filter (LVWF). A rainbow colour pattern is projected using the cylindrical lens and the LVWF. By detecting the colour of the pattern along the captured image of the camera, correspondence information between the camera and projector is obtained. The use of the LVWF allows for an infinite resolution for the projector, making...
spatial resolution to be solely dependent on the camera. Experimental results of scenes with a single object, uniform in colour, are presented. It’s shown that 2D texture information can also be obtain by either turning the projector off, or transforming the RGB colour image to Image System Hires (ISH) to obtain the gray-level intensity image. The main limitation of this technique is that the use of colour produces measurement errors when the pattern reflects from varying object surface colours [39].

2.2.1.3 Colour-encoded fringe projection technique

In [35], three phase-shifted sinusoidal fringe patterns with vertical fringes are coded into the three colour channels of the pattern projected by a Digital Light Processing (DLP) projector. A camera is used to sequentially capture each colour channel of the projected pattern and provide pixel to pixel correspondence between the camera and projector. Figure 6 shows the projected colour pattern onto a plaster head sculpture, the three captured images by the camera, and the 3D surface image obtained. The technique is advantageous in that a single pattern is loaded into the projector, while three images, one of each colour channel, are captured by the camera. This allows for fast acquisition of the three sinusoidal phase-shifted patterns. However, similar to the 3D rainbow camera, since colour patterns are used, the technique is susceptible to errors caused by varying object surface colours. This is evident in the rough surface of the plaster head sculpture shown in Figure 6(e).

![Figure 6: Colour-encoded fringe pattern: (a) Projected pattern, (b)-(d) Captured images of the three sinusoidal phase-shifted patterns (phase shifted by $2\pi/3$ rad), and (e) 3D surface image [35].](image)
The single pattern techniques discussed in these sections exhibit limitations that make them inefficient to use in environments where multiple objects with varying surface colours exist. The moiré technique uses fringes that defocus throughout a large working range; moiré surface contouring applications tune the hardware such that the patterns are focused with respect to a reference distance. The 3D rainbow camera and the colour-encoded fringe projection technique are prone to errors caused by varying object surface colours. Single colour encoded pattern techniques sacrifice accuracy for improved acquisition speeds. In general, accuracy of structured light systems can be improved by increasing the number of projected patterns [40]. This makes multiple coded pattern techniques a more suitable alternative.

2.2.2 Multiple Coded Pattern Techniques

Multiple coded pattern approaches project a distinct set of monochrome patterns and quickly switch between the patterns to minimize capture time.

2.2.2.1 Binary-coded technique

Binary-coded techniques code the pixels of the DLP projector with a pattern consisting of black and white intensities, representing 0 and 1 binary values. A series of binary patterns are projected and captured by a camera. For a given pixel in the captured images, its intensity values (0’s and 1’s) are used to determine its correspondence information with respect to the projected pattern. In [36, 37], a binary-coded technique that uses four patterns is presented. The system consists of an NTSC camera and a DLP projector. Each of the four patterns is captured at 60fps, resulting in a 15fps acquisition rate. The patterns projected on an elephant figurine are shown in Figure 7.

Figure 7: Binary-coded patterns [36].
The patterns contain black and white stripes with the property that each stripe’s boundary contains a unique code that can be decoded by comparing the four projected patterns. The limitation of this technique is spatial resolution. This is due to the fact that the stripes are required to be larger than one pixel in width, meaning that the correspondence information obtained is between the stripes and camera, rather than the individual pixels of the pattern and pixels of the camera. This results in a lower resolution of the projected pattern and therefore a lower spatial resolution. In [36], it has been determined that the error in 3D measurements depends on the accuracy of stripe boundary location. Although monochrome patterns are projected, stripe boundary identification is prone to errors caused by object texture.

2.2.2.2 Phase-shifting technique

The other type of multiple coded pattern techniques, the phase-shifting technique, utilizes a set of monochrome patterns to obtain pixel correspondence information between the camera and projector. In general, these techniques utilize a digital projector and a camera for pattern projection and capture, respectively. The use of a digital projector allows the projected patterns to be coded at pixel level using digital images sent to the projector via a computer. The use of monochrome patterns allows these techniques to obtain accurate 3D information of objects with varying surface colours. Furthermore, the patterns are coded at pixel level resolution, and therefore provide a higher spatial resolution than other multiple coded pattern techniques. These techniques all shift a set of fringes in the projected patterns to obtain pixel correspondence information, and vary by the type of fringes that are projected. Phase-shifting techniques include the sinusoidal fringe phase-shifting technique [38], the triangular phase-shifting technique [29, 41-44], and the trapezoidal phase-shifting technique [28, 39, 40]. The following section discusses the method of pattern generation for these types of phase-shifting methods and the techniques used to obtain pixel correspondence information for each.

2.3 Phase-Shifting Techniques

This section describes the codification of patterns used in various phase-shifting techniques and the method of decoding the patterns once they are projected and captured by a camera. Prior to the description of each individual phase-shifting technique, a general discussion on coding the projected images with phase values is provided.
The patterns to be projected are coded along one direction, for example the horizontal direction by phase values, \( \Psi(x,y) \), incrementally increasing from 0 to \( N \times 2\pi \), where \( x \) and \( y \) represent the horizontal and vertical axes of the image, and \( N \) is an integer representing the number of fringes that appear in the projected pattern. When the phase values are coded along the horizontal direction of the pattern, vertical fringes result, i.e. Figure 6 and Figure 7. The coded phase values represent the horizontal location of a given pixel in the projected pattern. In general, as \( N \) increases, the resolution of the projected fringe pattern increases. This is due to the fact that an 8-bit image can only be coded by integer values ranging from 0 to 255. The resolution of a projected fringe pattern can be described as follows:

\[
R_p = \frac{R_{N-h}}{\text{ceiling} \left( \frac{R_{N-h}}{N \times (I_{\text{max}} - I_{\text{min}})} \right)},
\]

where \( R_p \) is the resolution of the projected pattern, \( R_{N-h} \) is the native horizontal resolution of the projector, \( N \) is the number of fringes, and \( I_{\text{max}} \) and \( I_{\text{min}} \) are the maximum and minimum intensities of the projected pattern (ideally 255 and 0, respectively). Figure 8 shows the colour image of a rubble scene. If the phase values are directly coded into intensity levels by scaling them such that values 0 to \( N \times 2\pi \) range from 0 to 255, an image similar to the one shown in Figure 9 can be captured by the camera. The intensity profile of the projected pattern is shown above Figure 9.

![Figure 8: Rubble scene.](image)
2.3.1 Sinusoidal Phase-Shifting Technique

Sinusoidal phase-shifting technique utilizes a set of fringe patterns with sinusoidal intensity profiles. The monochrome fringe patterns are shifted by \( \frac{2\pi}{3} \text{rad} \) with respect to each other when projected. The patterns are then captured by a camera and the intensity values of each pixel in the captured images are used to identify its corresponding pixel in the projected pattern.
The phase values of the projected patterns are converted to fringe phase values, $\theta_p(i, j)$, ranging between 0 and $2\pi$, and used in Equations (2)-(4) to generate three sinusoidal phase-shifted patterns [50]:

$$I_1(i, j) = I'(i, j) + I''(i, j)\cos[\theta_p(i, j) - 2\pi/3]$$ (2)

$$I_2(i, j) = I'(i, j) + I''(i, j)\cos[\theta_p(i, j)]$$ (3)

$$I_3(i, j) = I'(i, j) + I''(i, j)\cos[\theta_p(i, j) + 2\pi/3]$$ (4)

For a given pixel $(i, j)$ in the projected image, where $i$ and $j$ represent the horizontal and vertical location of the pixel in the fringe, respectively, $I_1(i, j), I_2(i, j), \text{and} I_3(i, j)$ are the three captured intensity values, $I'(i, j)$ is the average intensity, $I''(i, j)$ is the intensity modulation, and $\theta_p(i, j)$ is the projected phase value that varies between 0 and $2\pi$ based on the position of the pixel in the fringe period. The resulting three patterns are projected onto the rubble scene and captured by a camera. The three captured sinusoidal phase-shifted patterns are shown in Figure 10. The intensity profile of each pattern is shown above its respective image.

![Figure 10: Sinusoidal phase-shifted fringe patterns: (a) Pattern 1, (b) Pattern 2, and (c) Pattern 3.](image)
To obtain pixel correspondence information between the camera and projector, the intensity levels of the three captured images of the projected patterns are used in Equation (5) [50]:

$$\theta_c(i, j) = \tan^{-1}\left[\frac{\sqrt{3} \cdot [I_1(i, j) - I_3(i, j)]}{[2 \cdot I_2(i, j) - I_1(i, j) - I_3(i, j)]}\right],$$

(5)

where $\theta_c(i, j)$ is the captured phase value referred hereon as the relative phase value for a given pixel $(i, j)$. When the relative phase values are directly coded into intensity values, a modulo $2\pi$ image is obtained, Figure 11. A modulo $2\pi$ image is an image whose relative phase values range from 0 to $2\pi$ for each fringe in the projected pattern, forming adjacent ramps with discontinuities in between. These relative phase values are insufficient for providing the correct camera and projector correspondence information as they only provide information about their respective fringe. The processing of obtaining the modulo $2\pi$ image from the captured images of the projected patterns is sometimes referred to as phase wrapping [35].

![Figure 11: Modulo 2\pi image using sinusoidal phase-shifted fringe patterns.](image)

Note the discontinuities outlined in red in the modulo 2\pi image. In order to obtain correct pixel correspondence information from the captured images, these discontinuities need to be removed using a phase-unwrapping technique [51] discussed in Section 2.4. Phase unwrapping techniques identify discontinuities in the modulo $2\pi$ image, and remove them by adjusting the relative phase values of fringes. The removal of discontinuities yields absolute phase values, $\theta'_c(x, y)$, that can be used to directly match with the coded patterns’ phase values, $\Psi(x, y)$, to obtain camera and projector pixel correspondence information.
Sinusoidal fringes provide an increase in projected pattern resolution. Since the intensity levels of the pixels within a sinusoidal fringe modulate six times between 0 and 255, a sinusoidal phase-shifted pattern with one sinusoidal fringe provides six times the resolution of a ramp pattern, such as the one shown in Figure 9. The modified version of Equation (1) becomes:

\[
R_p = R_{N-h} \left\lceil \frac{R_{N-h}}{N \times F_f \times (I_{\text{max}} - I_{\text{min}})} \right\rceil,
\]

(6)

where fringe factor, \(F_f\), is a constant based on the type of fringe used in the projected pattern. For a simple ramp fringe, \(F_f\) is one, while for the sinusoidal fringe, \(F_f\) is six.

In [38, 40], two methods of the sinusoidal phase-shifting technique are presented. The two techniques are similar in methods of pattern generation; however, differ in the way the modulo 2\(\pi\) image is obtained. The more traditional sinusoidal phase-shifting technique [38] is discussed here. The sinusoidal phase-shifted technique presented in [40] aims to improve the computation time of obtaining camera and projector correspondence information by replacing Equation (5) with a relatively more simple equation. This equation approximates the arctan function used in Equation (5), however introduces errors that require compensation in doing so.

The main limitation of the traditional technique is that it makes use of the arctan function to obtain pixel correspondence information. The use of the arctan is time consuming when compared with other structured light techniques and can limit acquisition speeds in applications where time is critical.

### 2.3.2 Triangular Phase-Shifting Technique

Triangular phase-shifting techniques use a set of fringe patterns with intensity profiles that form triangular shapes. The main motivation for using such patterns is to eliminate the need for the arctan function in obtaining the modulo 2\(\pi\) image and improve the technique’s computation time. The literature review presented here is of two research teams independently proposing a unique triangular phase-shifting technique. In [29, 41-44], triangular fringes depicting a sinusoidal fringe are presented. The fringe factor, \(F_f\), for the triangular fringes introduced is four. In [45-47], two triangular fringes are presented; one is a right-angled triangle fringe and the other is an isosceles triangle fringe, with \(F_f\) equal to one and two, respectively. A pattern with one right-
angled triangle fringe is simply the ramp pattern shown in Figure 9. The details of the technique proposed in [29, 41-44] is presented here.

In [29, 41, 42], a two-step triangular phase-shifting technique with (two triangular) patterns phase-shifted by $\pi$ is presented. To improve the accuracy of 3D measurements, the number of steps (number of phase-shifted patterns) is increased in [43, 44]. For simplicity of discussion, the two-step triangular phase-shifting technique is discussed here. The phase values of the projected pattern are used with Equations (7) and (8) to generate two triangular phase-shifted patterns [29]:

$$I_1(i, j) = \begin{cases} \frac{2I_m(i, j)}{T} i + I_{\text{min}}(i, j) + \frac{I_m(i, j)}{2}, & i \in \left[0, \frac{T}{4}\right] \\ \frac{-2I_m(i, j)}{T} i + I_{\text{min}}(i, j) + \frac{3I_m(i, j)}{2}, & i \in \left[\frac{T}{4}, \frac{3T}{4}\right] \\ \frac{2I_m(i, j)}{T} i + I_{\text{min}}(i, j) - \frac{3I_m(i, j)}{2}, & i \in \left[\frac{3T}{4}, T\right] \end{cases}$$

(7)

$$I_2(i, j) = \begin{cases} \frac{-2I_m(i, j)}{T} i + I_{\text{min}}(i, j) + \frac{I_m(i, j)}{2}, & i \in \left[0, \frac{T}{4}\right] \\ \frac{2I_m(i, j)}{T} i + I_{\text{min}}(i, j) - \frac{I_m(i, j)}{2}, & i \in \left[\frac{T}{4}, \frac{3T}{4}\right] \\ \frac{-2I_m(i, j)}{T} i + I_{\text{min}}(i, j) + \frac{5I_m(i, j)}{2}, & i \in \left[\frac{3T}{4}, T\right] \end{cases}$$

(8)

For a given pixel $(i, j)$ in the projected image, where $i$ and $j$ represent the horizontal and vertical location of the pixel in the fringe, respectively, $I_1(i, j)$, and $I_2(i, j)$ are the two intensity values of each pattern, $I_m(i, j)$ is the intensity modulation, $I_{\text{min}}(i, j)$ is the minimum projected intensity, and $T$ is the fringe pitch in pixels. The resulting two patterns are projected onto the rubble scene and captured by a camera. The two captured triangular phase-shifted patterns are shown in Figure 10. The intensity profile of each pattern is shown above its respective image.
The intensity values of the pixels in the two captured images are used in Equation (8) to obtain the modulo $2\pi$ image:

$$
r(i, j) = 2 \times \text{round} \left( \frac{R - 1}{2} \right) + (-1)^{R+1} \times \left\lfloor \frac{|I_1(i, j) - I_2(i, j)|}{I_m(i, j)} \right\rfloor.
$$

For a given pixel $(i, j)$, $R$ is an integer representing the region number of the pixel within the fringe period, $T$, and ranges from 1-4, $r(i, j)$ is the intensity ratio of the pixel within a given fringe and can be directly converted to $\theta'_c(i, j)$ by multiplying it into $2\pi$. Equation (8) yields a similar modulo $2\pi$ image as that shown in Figure 11. Similar to the sinusoidal phase-shifting technique, a phase-unwrapping algorithm can be used to obtain the absolute phase values, $\theta'_c(x, y)$, and yield the desired pixel correspondence information.

Both triangular phase-shifting techniques presented in this section are susceptible to error due to defocusing as the object distance to sensor changes from an optimal position [47]. In general, the more acute the interior angles of the triangle in its intensity profile are, the more significant the effect of defocusing becomes. Figure 13 shows the intensity profile deformation of the triangular fringe presented above when it is defocused. The dashed lines represent the defocused shape. As
the deformation increases, the fringe forms into a sinusoidal fringe. This phenomenon makes this technique unsuitable for applications that require a large working range for obtaining 3D information.

![Defocused fringe of two-step triangular phase-shifting method](image)

**Figure 13: Defocused fringe of two-step triangular phase-shifting method [29].**

### 2.3.3 Trapezoidal Phase-Shifting Technique

The trapezoidal phase-shifting technique uses a set of fringe patterns with trapezoid-shaped intensity profiles, Figure 14. Similar to the triangular phase-shifting techniques, the main motivation for using such patterns is to eliminate the need for the arctan function in obtaining the modulo $2\pi$ image and improve the technique’s computation time. This technique also attempts to address the defocusing issue that affects the accuracy of 3D measurements by using trapezoidal fringes. Trapezoidal fringes have larger interior angles than triangular fringes, and therefore are more robust to image defocusing. Furthermore, they provide higher resolution in the projected pattern with a fringe factor, $F_f$, of six. The method of trapezoidal phase-shifting pattern codification is discussed herein.
The phase values of the projected patterns are used with Equations (10) - (12) to generate three trapezoidal phase-shifted patterns [28]:

\[
I_1(i, j) = \begin{cases} 
I_0 + I^\prime, & \text{otherwise} \\
I_0(i, j) + I_0, i \in \left[ T \frac{6}{3}, \frac{T}{6} \right] \\
I_0(i, j) + I_0, i \in \left[ T \frac{2T}{3}, \frac{2T}{3} \right] \\
I^\prime(2 - \frac{6i}{T}) + I_0, i \in \left[ T \frac{T}{6} \right] 
\end{cases}
\]

(10)

\[
I_2(i, j) = \begin{cases} 
I_0(i, j), i \in \left[ 0, \frac{T}{6} \right] \\
I_0(i, j) + I^\prime(i, j), i \in \left[ T \frac{T}{6}, \frac{T}{2} \right] \\
I_0(i, j) + I^\prime(i, j), i \in \left[ T \frac{2T}{3} \right] \\
I_0(i, j), i \in \left[ 0, \frac{T}{3} \right] 
\end{cases}
\]

(11)

\[
I_3(i, j) = \begin{cases} 
I_0(i, j), i \in \left[ 0, \frac{T}{3} \right] \\
I_0(i, j) + I_0(i, j), i \in \left[ T \frac{T}{3}, \frac{T}{2} \right] \\
I_0(i, j) + I_0(i, j), i \in \left[ T \frac{5T}{6} \right] \\
I_0(i, j) + I_0(i, j), i \in \left[ 0, \frac{T}{6} \right] 
\end{cases}
\]

(12)

For a given pixel \((i, j)\) in the projected image, where \(i\) and \(j\) represent the horizontal and vertical axes of the image, respectively, \(I_1(i, j), I_2(i, j)\) and \(I_3(i, j)\) are the three intensity values of each pattern, \(I^\prime(i, j)\) is the intensity modulation, \(I_{\min}(i, j)\) is the minimum projected intensity, and \(T\) is the fringe pitch. The resulting three patterns are projected onto the rubble scene and captured by a camera. The three captured trapezoidal phase-shifted patterns are shown in Figure 14. The intensity profile of each pattern is shown above its respective image.
The intensity values of the pixels in the three captured images are used in Equation (13) to obtain the modulo $2\pi$ image:

$$r(i, j) = 2 \times \text{round} \left( \frac{R - 1}{2} \right) + (-1)^{R+1} \times \frac{I_{\text{med}}(i, j) - I_{\text{min}}(i, j)}{I_{\text{max}}(i, j) - I_{\text{min}}(i, j)}.$$  \hspace{1cm} (13)

For a given pixel $(i, j)$, $I_{\text{min}}(i, j)$, $I_{\text{med}}(i, j)$, and $I_{\text{max}}(i, j)$ are the minimum, median, and maximum intensities, respectively, $R$ is an integer representing the region number of the pixel within the fringe period, $T$, and ranges from 1-6, $r(i, j)$ is the intensity ratio of the pixel within a given fringe and can be directly converted to $\theta_c(i, j)$ by multiplying it into $2\pi$. Equation (7) yields a similar modulo $2\pi$ image as that shown in Figure 11. Similar to the sinusoidal phase-shifting technique, a phase-unwrapping algorithm can be used to obtain the absolute phase values, $\theta_c(x, y)$, and yield the desired pixel correspondence information.

In [28, 39, 40], the trapezoidal phase-shifting technique is presented. A faster computation time compared with the traditional sinusoidal phase-shifting technique is stated. However, it is also stated that even though the technique is more robust to defocusing effects when compared with triangular fringes, trapezoidal fringes are still susceptible to defocusing errors. Figure 15 shows the intensity profile deformation of the trapezoidal fringe when defocused. The dashed lines

---

**Figure 14: Trapezoidal phase-shifted fringe patterns: (a) Pattern 1, (b) Pattern 2, and (c) Pattern 3.**
show the deformed shape as the fringe defocuses. Similar to the triangular technique, the fringes start to take the shape of a sinusoidal fringe as it defocuses. This limitation makes this technique unsuitable for applications that require a large working range for obtaining 3D information.

![Defocused fringe of two-step triangular phase-shifting method](image)

**Figure 15:** Defocused fringe of two-step triangular phase-shifting method [28].

### 2.4 Phase Unwrapping

Phase unwrapping is the process of removing phase discontinuities that periodically appear in the modulo $2\pi$ image. Phase unwrapping is achieved by identifying the fringe number within which a given pixel lies, and adding the appropriate multiple of $2\pi$ to the pixel’s relative phase value to obtain its absolute phase value through the following relationship:

$$\theta_\epsilon(x, y) = (n(x, y) - 1) \times 2\pi.$$  \hspace{1cm} (14)

For a given $(x, y)$, $n(x, y)$ is the fringe number to which the pixel belongs. In high-quality modulo $2\pi$ images, this is a simple task. The phase discontinuity can be detected using simple conditional statements. However, in modulo $2\pi$ images that contain noise or are of objects with discontinuous or isolated surfaces, phase unwrapping can become more complex. The current techniques for phase unwrapping can be categorized into two categories: (i) path dependent and (ii) path independent techniques. The term “path” here refers to the order in which the pixels of the modulo $2\pi$ image are analysed and unwrapped. The following subsections discuss the two categories of phase unwrapping techniques in more detail.

#### 2.4.1 Path Dependent Techniques

Path dependent techniques determine the absolute phase of a given pixel by comparing its phase value with its neighbours. The path in which the pixels throughout the modulo $2\pi$ image are unwrapped depends on the technique used. Common path dependent phase unwrapping
techniques include: pixel queuing [52], tile processing [53], and the quality guided map [54-56]. In general, these techniques attempt to keep the affect of noisy pixels local, so that the error does not propagate to the rest of the image.

2.4.1.1 Pixel Queuing

The goal of the pixel queuing technique is to search for a path through the modulo $2\pi$ image such that the noisy pixels, and pixels that correspond to voids in the object surface, are unwrapped after the good quality pixels are identified and unwrapped [52]. The quality of each pixel is determined by comparing its relative phase value with its neighbour’s phase values. The quality of the pixel is inversely proportional to its phase difference with its neighbours. Figure 16 demonstrates the manner in which pixels are compared and an example of a path chosen by the algorithm.

![Figure 16: Pixel queuing: (a) Comparison of pixels, and (b) Unwrapping path [52].](image)

The limitation of this technique is that it can still produce incorrect results at points that correspond to voids in the object; however these bad points remain local due to the nature of the technique. Isolated surface regions cannot be unwrapped as a path between pixels corresponding to these regions and the rest of the image does not exist.
2.4.1.2 Tile Processing

In tile processing, the modulo $2\pi$ image is segmented into evenly sized squares (called tiles) [53], Figure 17. The relative phase values within the tiles are unwrapped and the tiles are stitched together using their intermittent phase values through the use of a Minimum Spanning Tree (MST) algorithm. MST’s objective is to determine an optimal path between the tiles so that error local to a tile does not propagate to the rest of the image through other tiles. This helps keep the effects of poor quality pixels local within a tile.

![Figure 17: Tiled section of phase unwrapped image [53].](image)

Propagation of errors within some tiles can still occur. The effectiveness of this technique is highly dependent on the tile size. The tile size needs to be adjusted depending on the object, making this technique unsuitable for applications that require phase unwrapping to occur without manual intervention.

2.4.1.3 Quality-Guided Map

Quality-guided map techniques are similar to pixel queuing in that they process the high quality pixels prior to poor quality pixels. The unwrapping path is determined by sorting the pixels from the highest to the lowest reliability values [54-56]. Figure 18 shows a numerical example of the
algorithm proposed in [54]. The first image represents the phase values of the pixels. The horizontal and vertical phase differences are shown by green and orange squares in image two, respectively. The propagation of colours (purple, blue, yellow and orange) through the pixel regions illustrates the phase unwrapping path chosen by the algorithm in this example.

![Quality-guided map algorithm’s path](image)

**Figure 18: Quality-guided map algorithm’s path [54].**

Quality-guided map techniques vary in the choice of reliability function and the design of the unwrapping path. In [54], the reliability of a pixel is determined by analysing the phase
difference of the pixel with respect to its horizontally, vertically, and diagonally adjacent pixels. In [55], a variation of the reliability function in [54] is used by only analysing horizontally and vertically adjacent pixels. In [56], a multilevel quality guided map technique is presented. The first level of quality map generation is achieved by removing the phase values that correspond to pixels in the background of an image. The second level of quality map generation is similar to the technique used in [55] to obtain the reliability of the remaining pixels.

The quality-guided map techniques are successful in unwrapping images with considerable amount of noise. Furthermore, surface voids of objects do not affect the accuracy of phase unwrapped pixels using this technique. However, regions that correspond to discontinuous surfaces (voids between objects) unwrap incorrectly as a path between them does not exist.

With added intelligence, path dependent algorithms can produce improved results in practice. However, pixel errors can remain undetected, causing the errors to propagate as the image is unwrapped. Furthermore, these techniques depend on the presence of a path between all pixels corresponding to object surfaces. In USAR environments however, a continuous surface between all objects cannot be guaranteed; standalone objects can leave isolated regions in the modulo $2\pi$ image without a path for these phase unwrapping techniques to follow.

### 2.4.2 Path Independent Techniques

Path independent techniques analyse the pixels of a modulo $2\pi$ image simultaneously using the relative phase values of their neighbours. Since all pixels are processed in parallel rather than sequentially, these techniques are known to be robust to noise [54]. However, processing the pixels in parallel requires a high number of iterations making such techniques computationally intensive. Common path dependent phase unwrapping techniques include: unwrapping by regions using least-squares approach [57], quad-tree decomposition [58], and regions referenced and window patching method [59].

In [57], phase unwrapping by regions using a least-squares approach is presented. At first, noisy pixels are removed from the phase wrapped image using thresholds. Then, the remaining phases are dividing into regions. Each region is unwrapped, and the regions are stitched together using a set of simultaneous linear equations. These equations describe the phase difference of pairs of pixels along the region boundary and are solved for the phase values of each pair using least-
squares. The technique is successful in obtaining an accurate phase unwrapped image even in the presence of noise and surface discontinuities; the effect of noisy pixels are kept from propagating. However, the technique’s success is highly dependent on the successful removal of noisy pixels, especially those that correspond to region boundaries. In addition, similar to other phase unwrapping techniques, the technique has limitations in dealing with completely isolated objects.

In [58], phase unwrapping by quad-tree decomposition is proposed, Figure 19. The modulo $2\pi$ image is segmented by checking if noise exists within a region (initially starting with one region comprising the entire image). The presence of noise in the region triggers the region to be sub-divided into four regions. Each region is then checked for noise and sub-divided again if noise is present. This process repeats until only noise free regions exist or regions have been divided down to a single pixel. Each region is then patched together using an error-norm minimization approach. Similar to all aforementioned techniques, the technique does not cope well with isolated surface areas due to the tile-combing method.

![Quad-tree decomposition method](image)

Figure 19: Quad-tree decomposition method [58].
In [59], a region-referenced window patching method is proposed for phase unwrapping. The modulo $2\pi$ image is segmented into tiles (referred in this technique as windows). The pixels within the window are unwrapped by considering the phase difference of a given pixel and its neighbours within a 5x5 square pixel boundary. The windows are then patched together in a predefined manner by considering the pixels corresponding to the window, and an overlapped region of its neighbouring windows, Figure 20. In Figure 20, windows are shown by the diagonal lines and the overlapped regions by the dashed lines. The technique performs well with noisy images. However, the accuracy of the results depends on the selection of the referenced-window shape and its threshold. Thus, this technique is not suitable for applications where manual intervention is not an option.

When compared with path dependent techniques, path independent techniques are inherently more robust to noise due to the fact that pixels are analysed in parallel, rather than in a sorted manner. In contrast, path independent techniques suffer from high computational intensities due to the large number of iterations required to obtain the unwrapped image. In general, however, both types of phase unwrapping techniques cannot accurately unwrap a modulo $2\pi$ image of a scene composed of multiple objects with isolated surface regions. The presence of clutter
consisting of multiple objects in USAR environments hinders the use of the aforementioned phase unwrapping techniques to provide accurate unwrapping of modulo $2\pi$ images. This necessitates the development of a new phase unwrapping technique capable of dealing with noisy pixels, as well as isolated regions of relative phase values.

### 2.5 Chapter Summary

Structured light techniques are classified into single and multiple coded patterns. Since single coded patterns tend to use colour in the projected image, their accuracy is limited in scenes with objects with various surface colours. Multiple coded patterns project a set of monochrome patterns to address this limitation. For USAR environments consisting of various types of objects with different surface colours, multiple coded patterns provide increased accuracy in obtaining 3D measurements. However, multiple coded techniques in addition to some single coded techniques, i.e. the colour-encoded fringe projection technique [35], require a phase unwrapping algorithm for obtaining camera and projector correspondence information. Phase unwrapping techniques, path dependents and path independent, have been developed to improve their immunity to noisy pixels. Path dependent techniques are less computationally intensive, however more prone to error propagation due to noise when compared with path independent techniques. Both types of phase unwrapping techniques lack the capability to correctly phase unwrap isolated surface regions of a scene. For the application of a structured light sensor in USAR environments, a multiple coded pattern technique using a phase unwrapping procedure capable of dealing with isolated surface regions is necessary.
Chapter 3
3D Sensory System

In this chapter, the development of the 3D sensory system for mapping USAR environments is presented. Namely, the hardware and software components of the sensory system are detailed, and the calibrations of the system parameters discussed. The overall 3D sensory system architecture is shown in Figure 21.

3.1 Hardware Development

The proposed 3D structured light sensory system uses a DLP projector, BenQ MP512ST, and a CCD camera, Prosilica GE680C, for pattern projection and acquisition, Figure 22. The DLP projector has a native resolution of 800x600 pixels and a brightness of 2200 ANSI lumens. The CCD camera has a resolution of 640x480 pixels. In summary, DLP projectors project images by passing white light through a rotating colour wheel. The colour wheel, composed of 5 segments
of colour, informs the projector of the activated channel using an encoder. The colour filtered light reflects off of a Digital Micromirror Device (DMD) chip and passes through the projector lens onto the objects in the scene. The DMD chip is composed of an array of micromirrors, each representing a pixel in the projected image. The micromirrors vibrate based on the intensity of the pixel; the intensity of the pixel and its micromirror vibration frequency are inversely proportional.

![Figure 22: 3D Sensory System Hardware.](image)

The DLP projector is modified using the technique described in [38] to remove the colour wheel and project the monochrome sinusoidal phase-shifted patterns at faster switching speeds. The encoder signal is replaced by a signal from a microcontroller unit, referred hereon as the trigger board. The trigger board signal is an impulse function with a period of 8.3ms and a pulse width of 0.2ms.

Due to the fact that only three of five colour segments of the colour wheel are unique (red, green, and blue) while the other two (yellow and white) are a combination of the first three channels, only three channels can be independently coded into an image and projected. Using a photodiode and oscilloscope, the interval of activation for each of the three channels is measured. Figure 23 shows the measured time interval of each channel of the projector, as well as the trigger board signal sent to the camera and projector. The projector’s projection time for one frame is
determined to be 8300\(\mu s\), or approximately 120fps. The green channel is active from 0-1760\(\mu s\), the red channel from 2200-3720\(\mu s\), and the blue channel from 6960-8240\(\mu s\). The yellow and white channels are between the red and blue channels, respectively.

Figure 23: Projector and Trigger board signal.

The trigger signal is sent to the CCD camera to trigger acquisition. The camera can be exposed for as little as 25\(\mu s\), however, it requires approximately 5ms of readout time (time required to digitize the CCD cell voltages). The camera can expose a frame while reading out another. However, a situation can arise when two frames are exposed during the readout of one frame. This occurs when the blue channel is in readout, while green and red are exposed. This results in loosing frames. Therefore, all three colour channels cannot be captured during one projection cycle. Instead, the green, blue, and red channels are captured during two cycles of acquisition, respectively. This allows enough time between the colour channels for the camera to readout its CCD chip information. The resulting acquisition frame rate of the three channels is 60fps.

### 3.2 Software Development

The 3D sensor is based on the multiple coded pattern structured light technique. In particular, it uses the sinusoidal phase-shifted pattern method of acquiring pixel correspondence information
between the camera and projector. Two sets of three sinusoidal phase-shifted patterns are used for 3D reconstruction; set 1 contains five sinusoidal phase-shifted fringes, while set 2 contains one sinusoidal phase-shifted fringe. Figure 24 shows captured images of the phase-shifted patterns of set 1 (a-c) and set 2 (d-f) with the intensity profiles of the images shown above their respective image. The first set of patterns is used to obtain the modulo $2\pi$ image, shown in Figure 11, and set 2 to use with a proposed phase unwrapping technique capable of dealing with isolated surface regions in the modulo $2\pi$ image.

![Figure 24: Captured images of projected patterns on the rubble scene: (a)-(c) Set 1, and (d)-(f) Set 2.](image-url)
The first set of patterns are coded into the colour channels of an image and sent to the projector to be projected and captured by the camera. During the acquisition of Set 1, the second set of patterns are coded into the colour channels of another image to be sent to the projector and captured by the camera. This results in an acquisition frame rate of 30fps (four projection cycles) for capturing the 6 patterns. This mode of capturing is referred to as the real-time view mode of the sensory system.

The captured images of the first set are used with Equation (4) to obtain the modulo $2\pi$ image shown in Figure 11. The following section describes the proposed active phase unwrapping technique used to unwrap the modulo $2\pi$ image to yield the absolute phase values corresponding to the projected image.

### 3.2.1 Active Phase Unwrapping

The issue with current phase unwrapping techniques discussed in Section 2.4 is that the absolute phase value for a given pixel depends on its relative phase value as well as the relative phase value of its adjacent pixels. Due to the presence of noise and surface discontinuities, these techniques are inherently prone to noise or error propagation. Furthermore, isolated surface regions cannot be correctly phase unwrapped as a relationship between their relative phase values does not exist. In such situations, an incorrect depth shift between the isolated surface regions results.

To obtain a correctly phase unwrapped image in instances where there are multiple objects, each pixel should be unwrapped independent of adjacent pixels. To do this, an additional set of sinusoidal phase-shifted fringe patterns is used, Set 2. These additional patterns contain a single fringe, which when wrapped provide one ramp in their modulo $2\pi$ image, Figure 25.
For a given pixel, its relative phase value in Set 2 can be used to determine its fringe number, $n$, and obtain its absolute phase value, $\theta_c(x, y)$, using Equation (15):

$$\theta_c(x, y) = \text{floor} \left( \frac{2\theta_c(x, y) \cdot N}{2\pi} \right) \cdot 2\pi + \theta_c(x, y).$$ (15)

For a given pixel $(x, y)$, $^1\theta_c(x, y)$ and $^2\theta_c(x, y)$ denote its relative phase values obtained from Set 1 and 2, respectively, and $N$ is the number of fringes in Set 1. The first term of Equation (15) obtains the fringe number of the pixel and determines its absolute phase value by adding $(n - 1) \times 2\pi$ to the second term (the relative phase value obtained from Set 1) of the equation.

Instead of comparing each pixel with its neighbours as is done in traditional approaches, the phase value of each pixel is independently obtained using the intensity values for each pixel in the six captured images. In the following subsection we present a comparison study to highlight the robustness of the proposed active phase unwrapping technique.

### 3.2.1.1 Active vs. Passive Phase Unwrapping Comparison

Presented herein is a comparison of the proposed active phase unwrapping approach versus a more traditional passive phase unwrapping technique, the quality guided map technique for the
scene presented in Figure 8. Figure 26 presents the absolute phase values of the objects in the rubble piles obtained using the two different approaches. Figure 26(a)-(c) show three consecutive phase unwrapped images using the active phase unwrapping technique, and Figure 26(d)-(f) show three consecutive phase unwrapped image using the quality guided-map technique. When accurate phase unwrapping takes place, the unwrapped phase image should reflect a gradual change in the intensity levels of the pixels from 0 to 255 moving from the left side to the right side of the image. Figure 26(a)-(c) show correctly phase unwrapped images of the scene as the variation in intensity is evident. For Figures 26(d)-(f), due to voids and discontinuities between objects in the captured scene, pixels corresponding to regions of discontinuity unwrap incorrectly using the passive phase unwrapping technique and this error propagates throughout the image to regions where objects are present and causes inaccuracies in the entire phase unwrapped image. When the image is unwrapped correctly, as seen in Figures 26(a)-(c), the unwrapped pixel intensity levels of the rubble on the right hand side of the image are close to white (between 200-255), indicating that these camera pixels correspond to the right side of the projected pattern. However, in Figures 26(d)-(f), these pixel intensities are closer to black (below 127 intensity level), indicating that the camera pixels correspond to the left side of the projected pattern. This error leads to inaccurate 3D information obtained from the scene. In addition, note the consistency in intensity levels of the phase unwrapped images obtained from the same scene throughout Figures 26(a)-(c). In Figures 26(d)-(f), however, these intensity levels vary in between the images. This demonstrates that the passive phase unwrapping technique provide inaccurate results. The active phase unwrapping approach does not analyse adjacent pixels corresponding to the pixel under consideration and therefore is able to provide accurate absolute phase values in cluttered environments.

Once the final phase unwrapped image is obtained using the proposed approach, the phase values of the pixels in the phase unwrapped image are matched with the projected pixels’ phase values in order to obtain the pixel-to-pixel correspondence between the camera and the projector. Triangulation is used to convert the correspondence information into 3D coordinate information using a phase-to-height algorithm similar to the one described in [60]. The 2D texture information of each pixel is obtained by averaging its intensity values in Set 1. Due to the one-to-one correspondence between the 3D coordinates and 2D texture information provided by the sensory system, the 2D texture information can be directly superimposed onto the 3D
coordinates to present the 3D textured coordinate information of the scene, Figure 27. Figures 27(a)-(c) show the 3D textured coordinate information obtained from the phase unwrapped images of Figures 26(a)-(c), and Figures 27(d)-(f) show the 3D textured coordinate information obtained from the phase unwrapped images of Figures 26(d)-(f). It is shown that not only does the quality-guided map approach incorrectly unwrap the relative phase values, its results are also non repeatable, making it difficult to intelligently account for its errors. Namely, this can be seen by the variation in the location of the isolated surface regions. By introducing the proposed active phase unwrapping technique, the 3D proposed sensory system is more robust to noise, surface discontinuities, and isolated surface regions, and therefore, able to provide a more accurate 3D map of a cluttered scene.

Figure 26: Phase unwrapped image: (a-c) active phase unwrapping technique, and (d-f) quality-guided map technique.
Figure 27: Textured 3D information obtained using: (a-c) the *active* phase unwrapping technique, and (d-f) the quality-guided map technique.

3.2.2 Shadow Detection

Occlusion refers to regions in a scene that are blocked from the field of view of one or more components in the 3D sensory system, Figure 28. Occluded regions of the projector in a structured light system are seen as shadows from the point of view of the camera. These shadow regions are referred hereon as cast shadows.

![Figure 28: Shadow and occluded regions.](image-url)
Due to the active nature of obtaining 3D measurements in the proposed 3D sensory system, cast shadows can result in inaccuracies in 3D reconstruction and map building of cluttered environments.

Currently, various methods for dealing with cast shadows in structured light applications have been proposed [61-65]. For example, in [61], three projectors and a camera have been used to project patterns onto cast shadow regions and obtain 3D information. In [62], additional optics, i.e. two collimating lenses and a half mirror have been used to manipulate the perspective of the projector to appear the same as the camera. The overlap of the two fields-of-view removes all occluded regions. In [63], mirrors have been used to bend the projected patterns around the object and onto the cast shadow regions. In [64], two cameras have been used to capture projected patterns; one camera provides information of regions occluded to the other camera. In [65], mirrors have been positioned behind the object to provide a full 360° view of the object to the projector and camera. These techniques all use additional hardware to deal with cast shadow regions by obtaining the 3D information belonging to these regions. The motivation for this is to obtain as much 3D information of the object as possible in one acquisition. However, since the proposed 3D sensory system is placed on a mobile platform, 3D information corresponding to cast shadow regions can be obtained when the robot moves the 3D sensory system. Namely, different viewpoints of the same objects will be provided. The cast shadow regions are identified and masked to minimize their effect on measurement errors. This concept will be further discussed in Chapter 4.

To detect pixels corresponding to cast shadow regions for the proposed 3D sensory system, only 3D sensory information within the working range of the sensor is considered. Since the light from the projector does not reach the shadow regions, the pixels corresponding to these regions maintain a constant intensity level within the captured images of the 2 sets of phase-shifted patterns. We utilize a pixel-wise comparison approach that analyses the 3 captured images Set 1 to determine the cast shadow regions. Namely, for a given pixel, if its intensity value remains constant in a tolerance bound within the 3 images, it is identified as a cast shadow pixel and therefore the pixel is masked. The tolerance bound is determined through colour calibration and is described in Section 3.3. Experiments on the performance of the shadow detection algorithm are provided in Chapter 4.
3.2.3 Capture Modes

As previously defined, the real-time view acquisition mode exposes each captured image to the scene for $1280-1760 \mu s$. Due to the short exposure time of the camera and the brightness of the projector, the optimal measurement range for this mode is 300mm to 800mm. This mode of acquisition is suitable for mobile robot navigation as it provides 3D information in real-time at 30fps. However, the range of 3D measurements is not sufficient for mapping. This requires the addition of a new acquisition mode.

To increase the range of 3D measurement, a new mode of capture, mapping capture mode, is defined. The camera exposure time is increased, resulting in a reduction of acquisition frame rate. Within this mode, two sub-modes of capture, short and long range capture are defined with $6960 \mu s$ and $12020 \mu s$ camera exposure times, respectively. For these sub-modes, the patterns can no longer be coded into the colour channels of the projector, as the channels are only exposed for a $1280-1760 \mu s$. The patterns for the short and long range sub-modes are coded as gray-scale images (all colour channels project the same intensity), and each pattern is sent to the projector as its own individual image. The increased exposure time can consequently saturate pixels corresponding to objects close to the sensor, and hence these sub-modes are utilized to capture objects within their range. The short range capture mode provides 300-1300mm range, while the long range capture mode provides 1300-2200mm range. By combining the 3D measurements obtained from the two capture modes, 3D information from 300-2200mm range can be obtained. Figure 29 illustrates the range that each mode of acquisition provides.

The two acquisition modes enable the operator to navigate the robot using real-time 3D information within 300-800mm range, and capture 3D information for mapping within 300-2200mm range. The following section discusses the system calibration procedure and results. Namely, colour calibration, and the estimation of the intrinsic and extrinsic parameters of the system are presented.
3.3 System Calibration

In this section, the calibration procedure for system colour calibration and the estimation of the intrinsic and extrinsic parameters of the system is discussed. In addition, the calibration results for each section are presented.

3.3.1 Colour Calibration

System calibration is the adjustment of the projected intensity levels of the projector such that the colour response curve of the camera is linear throughout its intensity response range. It is a necessary step in system calibration as it influences the working range and accuracy of 3D...
coordinate measurements. The colour calibration method is described in [38]. In summary, a flat white board is positioned in front of the 3D sensory system. The projection settings of the projector are adjusted such that its colour response curve is linear. The projector projects a series of patterns with constant intensity values throughout its intensity response range. The projected intensity levels and the intensity levels captured by the camera are recorded in a look-up-table (LUT). The projected versus captured intensity level information in the LUT are used to adjust the projected intensity levels such that the camera colour response range is linear. Due to the three acquisition modes of the 3D sensory system, this calibration procedure needs to be repeated for each mode. The live capture mode requires the calibration of the three colour channels of the projector, while the other two modes require the calibration of only one channel (referred hereon as the white channel), an entire image is captured rather than individual colour channels. The following subsections present the calibration results for each acquisition mode.

3.3.1.1 Real-Time View Acquisition Mode Colour Calibration

A linear curve is desired with respect to the intensities of the projected RGB colour channels and the captured intensities by the camera. In order to achieve this, the intensity for the colour channels was incrementally increased from the minimum to maximum intensity level of the linear response range of the projector. For the BenQ MP512ST, this linear response range is from 65 to 210. At each intensity increment, the intensity values of a 40x40 pixel area at the centre of the captured images were averaged. Figure 30(a) presents the intensity values prior to linearization. The intensity response of the camera for all three channels is from 30 to 185. The procedure after linearization is repeated and the intensity values after calibration are presented in Figure 30(b). The intensity response error after calibration for each channel is presented in Figure 31. Note that all estimated errors are within ±4 intensity levels. This tolerance is used in shadow detection as the tolerance for determining if a given pixel maintains a constant intensity level throughout the three captured images of Set 1.
Figure 30: Intensity response evaluation: (a) pre-calibration, and (b) post-calibration.
Figure 31: Linear intensity response error: (a) Red channel, and (b) Green channel.
3.3.1.2 Mapping Capture Mode

The following sections discuss the two sub-modes combined to obtain 3D coordinate measurements throughout the 300-2200mm range of the sensor.

**Short-Range Capture Mode Colour Calibration**

The same procedure implemented for the three colour channels in live capture mode colour calibration is repeated for the single (white) channel of the projector during short-range capture mode. The linear intensity response range of the projector is from 65-250. The intensity response of the camera is from 75 to 210. The colour responses of the camera pre- and post-calibration are presented in Figure 32. The intensity response error after calibration for the short-range white channel is presented in Figure 33.
Far-Range Capture Mode Colour Calibration

The same procedure is repeated for the white channel of the projector during far-range capture mode. The linear intensity response range of the projector is from 80-225. The intensity response of the camera is from 130 to 215. The intensity responses of the camera, pre- and post-calibration, are presented in Figure 32. The intensity response error after calibration for the short-range white channel is presented in Figure 33.

Figure 32(a): Short-range mode intensity response evaluation for pre-calibration.
Figure 32(b): Short-range mode intensity response evaluation for post-calibration.

Figure 33: Linear intensity response error
Figure 34: Short-range mode intensity response evaluation: (a) pre-calibration, and (b) post-calibration.
3.3.2 Intrinsic Parameter Estimation

Another important step in system calibration is the estimation of camera and projector lens’ intrinsic parameters. The lens intrinsic parameters provide a relationship between a point pixel in the image coordinate frame and the world coordinate frame of the same point in space (or in its field view). Precise estimation of intrinsic parameters is necessary to obtain accurate 3D coordinate information of objects using the 3D sensory system. A calibration method similar to [66] is implemented here. Two sets of sinusoidal phase shifted fringe patterns -one set of horizontal and one set of vertical fringes- are used obtain pixel to pixel correspondence between the camera and projector. This enables the images captured by the camera (referred herein as the camera image) to be transformed such that it appears as if the projector had captured the images (referred herein as the projector image), essentially transforming the projector into a camera. The camera calibration toolbox in Matlab by Bouguet [67] is used to obtain the intrinsic parameters of the camera lens using the camera images. Since the phase to height algorithm used to transform the absolute phase values obtained from the active phase unwrapping method to 3D
coordinate measurements uses a first order distortion model, the camera calibration toolbox in Matlab is constrained so that only the first order parameter is estimated. The same toolbox is used to calibrate the projector lens using the projector images. Since the two components are calibrated separately, this method separates the influence of calibration errors of one component onto the other.

A total of 93 images of a black and white checkerboard pattern positioned throughout the sensor’s range are captured to calibrate the camera and projector, respectively. The calibration results obtained from the camera calibration toolbox are presented in Appendix A.1. Using the pixel size of the camera and projector, 7.4µm and 16.6µm, respectively, the intrinsic parameter matrices obtained for the camera, $A^c$, and projector, $A^p$ are:

\[
A^c = \begin{bmatrix}
6.1346 & 0 & 2.3002 \\
0 & 6.1614 & 1.6977 \\
0 & 0 & 1
\end{bmatrix} \text{mm}, \text{ and}
\]

\[
A^p = \begin{bmatrix}
11.8261 & 0 & 9.7145 \\
0 & 11.6938 & 6.5300 \\
0 & 0 & 1
\end{bmatrix} \text{mm}.
\]

The error, defined as the difference between the coordinates of a checker corner point as computed from the real captured image and from the back projected image based on a nonlinear model, is shown in Figure 36. For the camera, the results show an error of +0.18 and -0.22 pixels in the x direction and +0.2 and -0.22 pixels in the y direction. For the projector, the results show an error of +1.25 and -1.4 pixels in the x direction and +1.25 and -1.6 in the y direction.

### 3.3.3 Extrinsic Parameter Estimation

Once the intrinsic parameters for both the camera and projector have been determined, the extrinsic parameters of the sensory system need to be obtained. The extrinsic parameters of the system describe the orientation and position of each component, the camera and projector, with respect to the world coordinate frame of the 3D sensory system, Figure 37. From Figure 37, the extrinsic parameter matrix describing the transformation between the camera and world coordinate frame, $M^c$, can be readily identified as $M^c$:
$M^c = \begin{bmatrix} -1.0000 & 0 & 0 & 0 \\ 0 & -1.0000 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ mm.

Figure 36: error caused by nonlinear image distortion for: (a) the camera, and (b) the projector.
For the estimation of the extrinsic parameters of the projector, a function called “Comp. Extrinsic” in the Matlab camera calibration toolbox is used. This function provides the transformation matrix between the capture device and a coordinate frame defined on the checkerboard pattern. The camera and projector images of the first orientation of the checkerboard pattern are used to define a common coordinate frame between the two components, Figure 38. Using the “Comp. Extrinsic” function of the camera calibration toolbox, the extrinsic parameters of each device with respect to the common coordinate frame is obtained, Appendix A.2. The two matrices along with $M^c$ are used to determine the extrinsic parameter matrix describing the transformation between the projector and world coordinate frames, $M^P$:

$$M^P =\begin{bmatrix} -0.9998 & 0.0013 & 0.0184 & -208.7438 \\ -0.0012 & -1.0000 & 0.0066 & -15.4538 \\ 0.0184 & 0 & 0.9998 & 8.1450 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{mm}.$$
3.4 Chapter Summary

A new 3D sensory system based on a structured light approach is proposed. It uses a multiple coded sinusoidal phase-shifting technique to obtain pixel correspondence information. Two sets of three sinusoidal phase-shifted patterns are used. The first set obtains the modulo $2\pi$ image, which is then phase unwrapped using the proposed novel active phase unwrapping technique. The active phase unwrapping technique utilizes the second set of three sinusoidal phase-shifted patterns to obtain the absolute phase values of the captured images, and determine the camera and projector pixel correspondence information. The use of additional patterns for phase unwrapping allows for the modulo $2\pi$ image to be correctly phase unwrapped without the propagation of noise and other errors due to discontinuous and isolated surface regions. A comparison between the proposed active phase unwrapping technique and a more traditional, quality-guided map, technique is presented. Furthermore, system calibration, namely the colour calibration and the estimation of system intrinsic and extrinsic parameters is discussed. Calibration results are presented with a discussion on error analysis.
Chapter 4
Experiments

Extensive experiments were conducted to verify the use of the proposed 3D sensory system for mapping USAR environments. These experiments focus on: (i) determining the sensor’s error in obtaining 3D coordinate measurements, (ii) analysing the performance of the sensor with various object compositions and surface profiles, (iii) effects of illumination conditions on 3D sensing, performance of shadow detection technique, and (iv) the 3D sensory system’s ability to provide sensory information for 3D mapping of USAR environments.

4.1 Measurement Error Estimation

To ensure that the 3D sensory system is capable of providing reliable 3D coordinate measurements for map building, a test environment consisting of a high-precision linear stage was utilized, Figure 39(a). The following subsections discuss the procedure for obtaining the measurement error in the z-axis, as well as the x- and y- axes of the 3D sensory system.

4.1.1 Depth (z-axis) Error

A USAR-like scene consisting of objects such as white drywall, concrete rubble, wood, metallic surfaces, and black tiles was created, Figure 40. A wooden flat board, Figure 39(b), was chosen as the reference surface as it provides the average amount of reflectivity with respect to the type of rubble-like objects that are within the USAR-like environment.

The board was attached to the linear stage in Figure 39 such that the sensory system’s optical axis (z-axis) is perpendicular to the board. Given that the two capture modes combined provide 3D coordinate measurements from 300mm to 2200mm of the sensor, the board was positioned throughout this range at 100mm increments. At each location, 20,000 measurement points were taken and used to determine measurement errors with respect to the true distance provided by the stage. These error values were then used to calculate the root mean square (rms) error of the sensor at each location, Figure 41.
Figure 39: Experimental setup: (a) overall set-up, (b) flat wooden board used in testing, and (c) checkerboard pattern used in testing.

Figure 40: USAR-like scene
As the object distance to sensor increases, the reduction of reflected projector light from the object back onto camera causes the deviation of measurement errors to increase. This decrease in depth measurement precision results in a maximum range of 2000mm for the 3D sensory system. In addition, it was determined that the minimum range of the sensor should be set to 500mm. Below this range, saturation of camera pixels due to the brightness of the reflected light occurs. Since the projected patterns modulate from high to low intensities, the saturation of low intensity pixels to high intensity causes sinusoidal waves to appear in a flat object. The effect of this error on 3D measurements is shown in Figure 42. The overall rms error for the optimal range of the sensor is determined to be 20mm.
4.1.2 X- and Y- axis Errors

A checkerboard pattern, with 35mm square sides, was attached to the board with its horizontal lines aligned with the x-axis of the sensor, and its vertical lines aligned with the sensor’s y-axis, Figure 39(c). The checkerboard was positioned throughout the range of the sensor (500-2000mm) at 100mm increments. At each increment, the x- and y- coordinate values of the checkerboard squares’ corners were recorded. By calculating the x-coordinate difference between horizontally adjacent corners, the measured distances of the squares’ horizontal sides were obtained. Similarly, the distances between vertically adjacent corners (vertical sides) were determined using the y-coordinate measurements. The error, defined as the difference between the measured horizontal and vertical square sides and the actual square sides, was determined. The x- and y- axis measurement errors were used to calculate the rms error of the sensor for each location. The rms errors with respect to checkerboard distance are shown in Figure 43(a) and (b), respectively. The rms errors are within 3.1mm throughout the sensor’s working range.

Figure 42: 3D point cloud of flat object at 350mm range. (Note the effects of saturated camera pixels).
Figure 43: Measurement error throughout sensor working range: (a) x-axis error, and (b) y-axis error.
4.2 Object Surface Measurement

For further analysis of the accuracy of 3D coordinate measurements obtained using the 3D sensory system, two objects (one rectangular prism, and one cylindrical) of known sizes were measured, Figure 44. The rectangular prism object was a gauge block with an actual measurement of 76.2mm, and the cylindrical object was a cylindrical weight with a diameter of 32.2mm (measured using a Vernier calliper). Each object was positioned in front of the 3D sensory system, and measurement results of a cross-section of their surface profile recorded. The measurement results obtained are shown in Figure 45. For the flat object, an average error of 0.87mm with a maximum error of 2.18 mm was determined. For the curved surface, an average error of 0.80 mm and a maximum error of 2.44mm were obtained. This demonstrates the 3D sensory system’s ability to obtain reliable results, regardless of the surface geometry of the objects.

Figure 44: Surface Measurement Objects: (a) 3” Gauge block, and (b) Cylindrical weight
Figure 45: Surface profile cross-section measurements of: (a) A flat surface (gauge block), and (b) Curved surface (cylindrical weight).
4.3 Object Surface Effects on Measurements

The accuracy and ability to obtain accurate 3D coordinate measurements of objects highly depends on their surface type. To evaluate the effects of object surfaces on 3D coordinate measurements, varying objects with different material compositions were tested. These objects include a small plastic fan, a block of concrete, an aluminum plate, a human face, a wooden board, and a mirror. The 2D texture and 3D coordinate images of each object are shown in Figure 46. In general, the 3D sensory system is robust in obtaining 3D coordinate measurements of objects with diffuse, matte or partially matte surface finishes.

The fan in Figure 46(a) is made of matte plastic and has a highly detailed propeller cover. The sensor is able to accurately obtain 3D information from this object including the details of the narrow slights in its cover. In addition, the partially shiny plastic propellers have also been accurately measured. The ability to obtain accurate measurements of human skin is important as the presence of victims in USAR environment is expected. The reflectivity of human skin allows for accurate 3D measurements as shown in Figure 46(b). Figure 46(c) presents the 3D coordinate measurements of a non-smooth concrete rubble block. The surface of the object allows for its details, i.e. the indents in the surface of the concrete, to be accurately obtained. Figure 46(e) shows the 3D coordinate measurements of a piece of wood. The surface type of this object also allows for accurate 3D coordinate measurements to be obtained using the 3D sensory system.

The 3D sensory system exhibits similar results to laser scanners and time-of-flight cameras when measuring objects with specular surfaces. This is due to the fact that the specular surface of an object reflects the projected light in an unpredictable manner, and at times saturates the CCD cells of the camera. This can be seen in Figure 46(d) and Figure 46(f). In Figure 46(d), a shiny aluminum plate is measured. It can be seen that the 3D coordinate information of a small slit on the plate’s surface is missing. This portion of the image is saturated in the camera images, such that all images show the same intensity level for the pixels belonging to this region. Due to the nature of the occlusion detection algorithm, this region is detected and masked. In Figure 46(f), a mirror with a wooded frame is measured. The glass portion of the mirror reflects the projected light in a direction away from the camera’s CCD array. As a result, only the wooden portion of the mirror (its frame) provides accurate 3D coordinate measurements.
4.4 Shadow Detection

As the robot traverses the USAR-like scene in Figure 40, its pose with respect to the objects in the scene is constantly changing, which may potentially cause a change in illumination as the 3D mapping sensor captures scene information, hence resulting in distinct shadows being casted. Figure 47 presents 2D and 3D sensory information taken at different robot poses with regards to the same objects in the USAR-like scene. The pixels identified within these shadow regions of the scene for the different robot poses are highlighted in blue in the 2D images. As can be seen in the images, the robot pose transformations introduced slight variations in illumination, which can be seen by the dissimilarity in the cast shadows identified in the scene. The masked pixels (pixels corresponding to shadow regions in the scene) for the different robot poses are highlighted in blue in the 2D images. In all cases, the shadow detection technique is able to identify and mask these shadow regions, hence minimizing their effect on the accuracy of the 3D sensory information.
Since the sensory system is an active system, the illumination conditions of the environment do not directly attribute to its sensing capabilities. In particular, the sensory system is capable of detecting objects within indoor lighting conditions, and dim lit and dark environments. A set of experiments were implemented to illustrate the sensory system’s robustness to different lighting conditions. Three lighting conditions utilized in the experiments are defined as: (i) dark, with all ambient indoor lighting turned off, (ii) indoor ambient lighting present, with all fluorescent lights
above the USAR-like scene on, and (iii) bright lighting conditions, consisting of using a closely
situated strobe light with a Xenon bulb near the objects in the scene. Figure 48 presents the 2D
texture image, 2D texture image with masked shadow pixel in cyan, and the 3D textured point
clouds of the objects obtained for each lighting condition. Note the consistency in the masked
shadow regions. This consistency illustrates the robustness of the shadow detection algorithm to
changing ambient lighting conditions. However, if the brightness of the ambient illumination
conditions continues to increase, the camera pixels can become saturated. In this case, the
shadow detection algorithm behaves similar to the case where an object’s surface saturates the
camera pixels; the pixels are detected as they maintain a constant intensity and masked by the
shadow detection algorithm. The loss of 3D information due to the high levels of brightness in
illumination conditions is common to all 3D sensors.

Figure 48: 2D texture, 2D texture with masked shadow pixels, and 3D textured point cloud
information of objects with various ambient lighting conditions: (a) dark condition, (b)
ambient fluorescent room lighting, and (c) bright lighting condition using strobe light.
4.6 3D Mapping

The ability to effectively use the proposed 3D sensory system for mapping USAR environments is of primary interest in this thesis work. The 3D sensory system was mounted on the robot platform to traverse throughout the USAR-like scene in Figure 40. As the robot traversed the scene, 3D sensory information, as well as 2D texture information of the rubble was acquired. In order to ensure overlap between sensory information, sensory data was taken at every 5-25cm of robot travel. An appropriate location was determined based on the overlap that the field-of-view of the 3D sensory system provided with respect to the field-of-view of the previous image taken. A 30-50% overlap was desired. The 3D coordinate frame of the first acquisition is defined as the global coordinate frame of the 3D map. For each subsequent acquisition of 3D information, its previous acquisition is used to stitch the measurements onto the global map.

In order to stitch two consecutive sets of 3D data, a Scale Invariant Feature Transform (SIFT) clustering technique [68, 69] is utilized to identify a distinct set of features in the corresponding 2D texture image of each acquisition. The SIFT technique implements a series of image manipulations based on pixel intensity distributions to obtain feature keypoints, each represented by 128 descriptors. The high number of descriptors allows the identified keypoints to be robust to image rotation, scale and illumination change when matching them between a pair of 2D texture images. Furthermore, an order of several thousands of keypoints can typically be extracted from a single image, resulting in a high number of matched keypoints using the keypoint descriptors. Figure 49 shows the identified keypoints in the first pair of 2D texture images. The keypoints are highlighted in green. Keypoints identified in the images are matched with keypoints in consecutive images using their descriptors. Figure 50 shows the matched keypoints of the first pair of 2D texture images obtained from the scene. Matched keypoints in the two 2D texture images are connected by blue lines.
Since there is a one-to-one correspondence between the 2D texture image and the 3D coordinate measurements, the pixel image coordinate of the matched keypoints can be used to identify their 3D coordinate measurements. The coordinate information of the matched keypoints are used in an Iterative Closest Point (ICP) algorithm [70] to register the pair of 3D textured point clouds into a global coordinate frame. This is done by minimizing the sum of squared differences between pixels and their closes neighbours, yielding a local minimum solution. The success of this technique is dependent on the initialization of the point clouds to help find the correct local minimum solution. The registration of the point clouds into a common global coordinate frame is performed using the transformation $(R, t)$ matrices obtained by minimizing Equation (16):
\[
E(R,t) = \sum_{i=1}^{N_k} \sum_{j=1}^{N_l} w_{i,j} \left\| P_i - (RP_j + t) \right\|^2,
\]

where \( N_k \) and \( N_l \) are the number of points in point sets \( P_k \) and \( P_l \), and \( w_{i,j} \) is the weights for a point match. When \( P_i \) is the closest point to \( P_j \) within a defined bound, \( w_{i,j} = 1 \), otherwise, \( w_{i,j} = 0 \). The transformation matrix, \( ^1T_2 \), obtained describes the transformation of the coordinate frame of the second point cloud with respect to the coordinate frame of the first point cloud. The same procedure is repeated for the third point cloud to obtain \( ^2T_3 \). Since the coordinate frame of the first point cloud is used as the global coordinate frame of the 3D map, the 3D coordinate information of the third point cloud is translated onto the global coordinate frame using the transformation matrix \( ^1T_3 = ^1T_2 \times ^2T_3 \). Figure 51 shows the first three 3D point clouds stitched together using the described procedure.

Figure 51: Stitching of the first three 3D point clouds.

The following subsections discuss the mapping results as well as presenting an evaluation of the capability of the sensor in providing reliable information for feature (keypoint) detection and matching.
4.6.1 3D Map Building Results

The entire scene required a total of 131 3D point clouds to build the map. An average 227,756 3D coordinate points were obtained per acquisition by the 3D sensory system. Figure 52 shows the overall 3D map with superimposed texture information of the entire scene. In Figure 52(a), the global coordinate frame of the 3D map is shown in red. The robot’s path is shown by the green line in Figure 52(a) and is determined by connecting the coordinate frame origins of each 3D point clouds in the order of capture. The path begins with a dark shade of green for the first 3D texture image, and becomes brighter as the acquisitions progresses. In Figure 52(b), a different perspective of the 3D map is presented. Zoom-in views of various locations, outlined in red, of the 3D map are provided in Figure 53. Zoom-in 1 shows a textured point cloud of a Barbie doll representing a victim with the majority of the doll being covered by concrete rubble. Zoom-in 2 shows another doll representing a victim. Zoom-ins 3 to 7 show the concrete rubble piles of the scene at various locations. Zoom-in location 4 shows two isolated surface regions of objects within the scene. Zoom-in 8 shows one side of the ramp close to the centre of the scene. Zoom-in 9 shows a mannequin head representing another victim located under the ramp. The zoom-in views demonstrate the ability of looking at various objects in a 3D map from different perspectives, aiding the operator in object and victim recognition.

4.6.2 Recall Rate

An important aspect of 3D visual mapping is the performance of the feature detection algorithm used. To evaluate the 3D sensory system’s performance in providing reliable information for feature detection in cluttered scenes, the recall rate of the SIFT algorithm is evaluated. A similar technique as that proposed in [71] is presented. Herein, the recall rate for this application is defined as the ratio of matched keypoints between a pair of images and the total number of keypoints identified in the overlapped region of the pair.

For the 131 3D point clouds obtained, the recall rate as defined is obtained with respect to the change in the 3D sensor’s optical axis as the robot traverses the scene. Figure 54 shows the plot of the recall rate estimation with respect to the translation and rotation of the 3D sensory system’s optical axis.
Figure 52(a): 3D map with superimposed robot path.
Figure 52(b): Alternate view of 3D map.
Figure 53: Zoom-in views 3D map.
The distribution of the results is consistent with the results presented in [71]. Figure 54 shows that when the displacement is small (less than 100mm) and rotation is large (10°-15°), the recall ranges between 0.35-0.75. On the contrary, when the displacement is large (between 100-200mm) and rotation is small (less than 5°), the recall rate ranges between 0.1-0.25. This indicates that the displacement of the 3D sensory system’s optical axis has a larger influence on recall rate than its rotation. The large drop in recall rate with large optical axis displacements is due to the movement of the dominant light source (the projector) between the 3D point cloud pairs. Based on the results of the experiments performed in this thesis work, the SIFT technique is robust to ambient illumination changes, however, less robust to variations in the location of the illumination source. The translation of the light source causes fewer matches to be found between a pair of images, thus a lower recall rate. The rotation of the light source has less influence on the distribution of pixel intensities across the image. Consider a SIFT keypoint (shown in red) in Figure 55(a). The pixel intensity distribution of this keypoint is based on the
light ray from the projector incident on this point (shown in green). When the 3D sensory system’s optical axis rotates, a different light ray from the projector is incident on this keypoint. However, the angle of incidence of the light ray on the keypoint remains the same. On the contrary, the angle of incidence of the light ray from the projector onto the keypoint changes as the optical axis of the 3D sensory system translates. This causes a more drastic change of pixel intensity distributions near this keypoint, resulting in the change of its descriptors.

![Figure 55: Pose change influence on SIFT keypoint descriptors: (a) Identified SIFT keypoint, (b) Rotation of optical axis, and (c) Translation of optical axis.](image)

In this work, a recall rate above 0.4 is desired. A recall rate above 0.4 is achieved when the displacement values are less than 100mm in combination with rotation angles between 5°-15°. Lower recall rates (below 0.1) result from displacements greater than 240mm in combination with rotation angles greater than 15°. These results indicate that in order to achieve a good performance from the feature detection algorithm, the relative pose change of the 3D sensory system should be limited to less than 200mm translation and 15° rotation.

### 4.7 Discussion of Experimental Results

The error estimation results in Section 4.1 show an rms error of 20mm for the sensor’s depth measurements, and 3mm in the x- and y- axes of the 3D sensory system’s coordinate frame. The increase in measurement errors as the object distance to sensor increases can be attributed to the reduction in the intensity of the reflected light from the object. Two factors contribute to this reduction: (i) the geometry of the DMD and CCD chips of the projector and camera, and (ii) the
diffraction of light rays from the projected pixels. Between each pair of pixels on the components chips (CCD or DMD), there exists a gap shown in Figure 56. In this figure, the gray chip represents the DMD chip, the purple chip represents the CCD chip, and the white squares inside the chips represent the pixels for each component. As shown in Figure 56, the alignment of the CCD chip’s pixels with respect to the DMD chip’s pixels can influence the amount of captured intensity levels. When the CCD pixels are exposed to 90% of the DMD pixels, Figure 56(a), a considerable amount of reflected light is captured by the camera. However, as the two components align such that the CCD pixels are exposed to less of an area covered by the DMD pixels, Figure 56(b), the reduction in reflected light intensities is increased. Also, consider a zoom in view of a portion of the projected pattern shown in Figure 57. When the object is close to the sensor, the specular reflection of the light rays causes the pixel to appear as a circle larger than its actual size. This is shown by the dashed circle on pixel $(i, j)$ in Figure 57(a). This reduces the effects of reduced reflected light caused by gaps between the CCD and DMD chip pixels. However, as the distance of the object to the sensor increases, the diffraction of light rays causes less light to reach the object. Consequently, the pixels appear as a circle smaller than its actual size, Figure 57(b). This not only causes a reduction in the intensity of the reflected light, it also contributes to the effects of factor (i) as mentioned above. Nonetheless, the estimated measurement errors are comparable with measurement errors of other 3D sensors intended for USAR environments such as the 3D laser range finder developed in [18] which provides an error bound of 26 mm up to its maximum range of 300mm and the 3D TOF camera used in [21] which provides an error bound of approximately 110mm up to 2000mm range.

The results of the experiments for measurement error with respect to the surface profile of the object are consistent with the measurement errors obtained with respect to the 3D sensor’s coordinate axes. The 3D sensor is shown to be able to obtain accurate 3D coordinate information, regardless of the geometry of the surface of the object. The results for the cast shadow detection experiment showed that the 3D sensor is able to detect pixels corresponding to shadow regions. The algorithm’s robustness to illumination changes is also presented in a latter section, Section 4.5. It is shown that the algorithm is consistent in identifying cast shadow regions in various indoor illumination conditions.
Figure 56: CCD and DMD chip geometric representation: (a) Less influential alignment, and (b) More influential alignment.

Figure 57: Illumination reduction illustration. Object distance is: (a) close to the sensor, and (b) far from the sensor.

The results for analysing the effects of object surface types show that the proposed 3D sensory system experiences similar challenges as other active 3D sensors in obtaining information from specular surfaces. Since specular surfaces saturate the camera pixels corresponding to these regions. As a result, the occlusion detection algorithm detects regions corresponding to specular surfaces, and masks their pixels accordingly.
The ability to obtain accurate 3D point clouds is shown through experiments with varying illumination conditions. The 3D coordinate measurements are shown to be unaffected by illumination changes. However, as the ambient illumination condition become dimmer, the brightness of the 2D texture image decreases as well. Although this does not affect the accuracy of the 3D coordinate information, it can make detecting objects and victims difficult for a robot operator. By adjusting the pixel intensity levels of the 2D texture image so that their brightness increases, a brighter 2D texture image can be provided to the operator. Adjusting the pixel intensity levels can easily be automated by analysis the intensity levels of the 2D texture image; if the intensity levels fall below a threshold, the brightness of the 2D texture image is increased.

The results of 3D mapping show promise for the use of the proposed 3D sensory system for mapping USAR environments. The accuracy of the 3D information, the sensor’s robustness to various shapes, surface types and colours, and illumination changes show its capability in providing accurate results for this application. The estimation of recall rates with respect to the pose change of the 3D sensory system show that the 3D information obtained from the sensory system can reliably be used with feature detection algorithms. The 3D sensor is shown to perform well with the mapping algorithm utilized.

4.8 Chapter Summary

In this chapter, an extensive analysis of the performance of the proposed 3D sensory system is presented. Experiments were conducted to evaluate the measurement error of the 3D sensor with respect to each of its three coordinate axes. Further evaluation of the measurement ability of the sensor is provided by scanning a flat and a curved object. In addition, the performance of the proposed occlusion detection algorithm is presented by changing the 3D sensory system’s pose with respect to a static scene. The masked pixels detected as pixels corresponding to occluded regions are shown for each pose. The algorithm’s robustness to illumination changes is also presented in a latter section.

The ability to obtain accurate 3D point clouds is shown through experiments with objects of various surface types, and environments of varying illumination conditions. The 3D measurements are shown to be unaffected by illumination changes. However, object surface types that are specular, such as a mirror, or shiny aluminum plate, are shown to yield regions of
missing 3D information. Nonetheless, the remaining regions of 3D information provide accurate results and therefore can be used for 3D mapping.

Finally, a 3D map of the USAR-like scene demonstrates the proposed 3D sensory system’s ability in obtaining 3D information of such environments. The sensory system is shown to provide reliable results for feature detection that is used for stitching 3D point cloud in order to build the map. Zoom-in views of the map further illustrate the performance of the 3D sensory system in mapping USAR environments.

Overall, the experimental results presented in this chapter show promise for the use of the proposed 3D sensory system for mapping of unknown cluttered USAR environments. The accuracy of the 3D information, the sensor’s robustness to various shapes, surface types and colours, and illumination changes show its capability in providing accurate sensory information for this application.
Chapter 5
Conclusion

Due to the limitations of current 3D sensors for robot exploration and mapping in USAR environments, a new 3D sensory system based on a structured light technique is proposed in this thesis. The 3D sensory system utilizes an active light source to obtain 3D information, as well as 2D texture information of cluttered unknown environment. Two modes of capture (real-time view and mapping) are presented. The real-time view mode provides 3D measurements from 300mm to 800mm range at 30fps, suitable for robot teleoperation through a cluttered scene. The mapping capture mode provides 3D measurements from 500mm to 2000mm range, making this mode suitable for obtaining 3D measurements used for map building.

5.1 Summary of Contributions

The primary contributions of this work are summarized below:

5.1.1 A 3D Structured Light Sensor for USAR Missions

In cluttered scenes common to USAR environments, the presence of various objects with different surface types and colours, as well as the varying illumination conditions provide challenges for current 3D sensors proposed. Passive stereo vision sensors cannot obtain 3D information in dimly lit or dark environments, 3D laser range finders cannot provide 2D texture information, and 3D TOF cameras provide 3D measurement with low resolution. Due to such limitations, a new 3D sensory system based on structured light is proposed. Since current structured light techniques are only used for obtaining 3D measurements of a single stand-alone object, a new structured light technique is proposed to address this limitation. The contributions of this work in developing a 3D sensory system for USAR missions are as follows.

5.1.1.1 Hardware Development

A 3D sensory system based on structured light was developed for robot exploration and mapping in USAR environments. The sensor consists of a DLP projector and a CCD camera synchronized for pattern projection and image capture. The colour wheel of the projector was removed to enable pattern projection and capture in monochrome (gray-scale) at 30fps within a working
range of 300mm to 800mm. The exposure time of captured images by the camera was increased to provide a working range of 500mm-2000mm for obtaining 3D information.

5.1.1.2 Software Development

Namely, a new active phase unwrapping technique has been proposed within the structured light framework to address the limitations of current structured light techniques. It enables the 3D sensory system to provide reliable 3D measurements in cluttered environments consisting of multiple objects with various surface shapes and textures. It also deals with isolated surface regions.

In addition, a cast shadow detection method for structured light techniques is developed to remove unreliable 3D measurements corresponding to shadow regions. This is mainly due to the fact that the structure light sensor relies on projected patterns for obtaining 3D information. When cast shadow regions are present due to the cluttered nature of the environment, the fringe patterns cannot be projected onto these regions. As a result, 3D measurements obtained from these regions are inaccurate. The proposed shadow detection method detects pixels corresponding to such regions and removes them from the collected data set of 3D information.

5.1.2 Implementation

The proposed 3D sensory system was mounted on a rugged rescue robot platform. A USAR-like test scene consisting of various materials common to USAR environments was constructed. Experiments with varying ambient illumination conditions and objects of various surface types were conducted. The 3D sensory system is shown to provide reliable measurements of varying object surfaces and shapes. The experiments in varying ambient illumination conditions demonstrate the sensor’s ability to obtain reliable information from dark, dimly lit, and bright illumination conditions. The performance of the shadow detection algorithm was verified based on changing robot poses. 3D point clouds obtained from the entire USAR-like scene were used to build a 3D map. The reconstructed 3D map of the scene with texture information demonstrates the accuracy of measurements obtained from the 3D sensory system.
5.2 Discussion of Future Work

The proposed 3D sensory system is an improvement over current 3D sensors intended for rescue robot navigation and mapping applications. The 3D sensory system’s ability to provide 3D coordinate measurements, as well as 2D textured images with one-to-one pixel correspondence makes this sensor especially suitable for 3D mapping applications for improved accuracy and computation times. Nonetheless, addressing some issues may strengthen the performance of the proposed 3D sensory system.

For robot navigation in the USAR-like scene, the real-time view mode of the 3D sensory system provides a sufficient measurement range. However, to use the 3D sensory system in less cluttered environments where the objects are further apart in the z-axis, the increase in range for this mode of acquisition can be an asset. To increase the range of the real-time view mode, a projector with a higher brightness can be used. This will enable the maximum range of the 3D information obtained during real-time view mode to be increased.

Another issue that can improve the accuracy of 3D coordinate measurements provided by the 3D sensory system is the minimization of the effects of reduced reflected light from the objects onto the camera’s CCD chip. As previously mentioned, two factors contribute to this: (i) the gap between pixels on the DMD and CCD chips of the projector and camera, and (ii) the diffraction of a pixel’s light rays over distance. The diffraction of pixels is analogous to the expanding field-of-view of the projector. The reduction of the effects of this factor can be achieved by choosing a projector with a longer focal length. Consequently, this results in a smaller field-of-view, and therefore a smaller area of 3D measurements per acquisition. This can significantly increase the number of 3D measurements required to map an area and is not desired. The second factor that contributes to the reduction of reflect light can however be improved. Selecting a projector with a higher DMD chip resolution (native resolution) minimizes the size of gaps between each pair of pixels on the DMD chip. This reduction helps the CCD chips’ pixels to be exposed to a larger portion of the DMD chip, and therefore capture more of its reflected light from the object.

The aforementioned future work can provide an improvement in 3D sensory information obtained from the proposed 3D sensory system. The increase in measurement range of the real-time view mode provided by a brighter projector can help navigation in larger and less cluttered
environments. A higher resolution projector can minimize measurement errors at longer ranges, hereby increasing the maximum range of mapping capture mode.

5.3 Final Concluding Statement

Overall, the proposed 3D sensory system can be a great asset to rescue robots used for navigating and mapping USAR disaster scenes. The proposed structured light-based sensor is the first of its kind for obtaining reliable 3D measurements of various objects in cluttered scenes. It provides two modes of acquisition: real-time view mode for teleoperation of the robot in scene exploration and mapping capture mode for 3D mapping of the environment. The ability to capture 3D sensory information and 2D texture images simultaneously using only two components (a camera and a projector) enables such a system to be designed in a compact package, allowing for easy integration onto mobile rescue robot platforms. Proposed hardware improvements can potentially increase the working range of the 3D sensory system for application in USAR environments.
References


[50] H. Schreiber and J. H. Bruning, “Phase Shifting Interferometry,” in Optical Shop Testing,


O. Wulf, A. Nüchter, J. Hertzberg, and B. Wagner, “Benchmarking Urban Six-Degree-of-
Appendices

A.1 Intrinsic Parameter Estimation Results

A.1.1 Camera Intrinsic Parameters

%-- Focal length:
f_c = [ 829.819320653191200 ; 832.623228176467710 ];

%-- Principal point:
c_c = [ 310.845662841962680 ; 229.414318518465960 ];

%-- Skew coefficient:
alpha_c = 0.000000000000000;

%-- Distortion coefficients:
k_c = [ -0.205478203563393 ; 0.000000000000000 ; -0.000000000000000 ; -0.000000000000000 ; 0.000000000000000 ];

%-- Focal length uncertainty:
f_c_error = [ 3.154100492038667 ; 3.149479403613869 ];

%-- Principal point uncertainty:
c_c_error = [ 0.981106130841100 ; 1.090782470637543 ];

%-- Skew coefficient uncertainty:
alpha_c_error = 0.000000000000000;

%-- Distortion coefficients uncertainty:
k_c_error = [ 0.002136533640335 ; 0.000000000000000 ; 0.000000000000000 ; 0.000000000000000 ; 0.000000000000000 ];

A.1.2 Projector Intrinsic Parameters

%-- Focal length:
f_c = [ 739.131086995627360 ; 730.864173860954450 ];

%-- Principal point:
c_c = [ 607.158472941794800 ; 408.126199625268270 ];

%-- Skew coefficient:
alpha_c = 0.000000000000000;

%-- Distortion coefficients:
k_c = [ 0.021886762213349 ; -0.000000000000000 ; -0.000000000000000 ; 0.000000000000000 ; 0.000000000000000 ];
%-- Focal length uncertainty:
fc_error = [ 10.593126162830529 ; 10.333722965628501 ];

%-- Principal point uncertainty:
cc_error = [ 0.000000000000000 ; 0.000000000000000 ];

%-- Skew coefficient uncertainty:
alpha_c_error = 0.000000000000000;

%-- Distortion coefficients uncertainty:
kc_error = [ 0.002016322769775 ; 0.000000000000000 ; 0.000000000000000 ; 0.000000000000000 ; 0.000000000000000 ];

A.2 Extrinsic Parameter Estimation Results

A.2.1 Camera-Common Frame Transformation

\[
\begin{bmatrix}
-0.01648925 & 0.99961102 & -0.02249212 & -262.3979194 \\
0.99985921 & 0.01641505 & 0.00347932 & -71.63512162 \\
-0.00310876 & 0.02254633 & -0.99974096 & 490.81269365 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\( \text{cam}_T_{com} \)

A.2.2 Projector-Common Frame Transformation

\[
\begin{bmatrix}
-0.01527622 & 0.99987510 & -0.00405052 & 62.61327032 \\
0.99987808 & 0.01528911 & 0.00317294 & -59.30939134 \\
0.00323447 & 0.00400156 & -0.99998676 & 481.21221103 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\( \text{proj}_T_{com} \)