**ADVANCEMENTS TOWARD AUTONOMOUS ACCESS HOLE DETECTION FOR
URBAN SEARCH AND RESCUE**

By
Ben Waismark
Honours Bachelor of Computer Science,
University of Waterloo 2011

A Thesis
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Science
in the program of
Computer Science

Toronto, Ontario, Canada
© Ben Waismark 2017

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Advancements Toward Autonomous Access Hole Detection For Urban Search And Rescue

Master of Science 2017

Ben Waismark

Computer Science

Ryerson University

ABSTRACT

The collapse of buildings creates voids underneath rubble, where victims may be trapped. First responder crews arriving at a collapse scene are responsible for the location of access holes, among other tasks. Access holes are entry points through which rubble voids may be reached by search and rescue personnel. Previously presented work aimed to autonomously locate such holes, aiding concentration of resources to areas of interest, such as those leading into rubble.

The work proposed improves upon existing work by increasing accuracy while reducing total number of detections. A new process is introduced for segmentation of colour and depth images, significantly improving the hole finding system's accuracy. The ability to score holes based on multiple frames, considering various points of view is introduced. As an added benefit, matching holes across frames allows the algorithm to report holes once per group of matches, rather than once per frame.

ACKNOWLEDGEMENTS

First, I wish to sincerely thank my supervisor, Dr. Alexander Ferworn, for his support, direction and encouragement throughout my studies.

I would also like to thank all members of the N-CART lab for providing me with a skilled, friendly and motivating environment. A special thank you to Jimmy Tran who has provided me with valuable knowledge and advice during my time at the N-CART lab.

Finally, to my wife, Asya Waismark, thank you for the love, support and encouragement you have given me along the way.

TABLE OF CONTENTS

Author's Declaration.....	iii
Abstract	iv
Table Of Contents	vii
List of Tables	x
List of Figures	xi
Abbreviations	xii
Introduction.....	1
1.1 Objective	4
1.2 Contributions	5
1.3 Organization	6
Related Work	8
2.1 Urban Search and Rescue (USAR)	9
2.1.1 Rubble And Collapse Survival	9
2.1.2 Signs Of Life	11
2.2 Data Acquisition.....	12
2.2.1 Sensors.....	13
2.2.2 Ground-based Robotics	15
2.2.3 Aerial Robotics	17
2.2.4 Disaster Dogs.....	18
2.3 Visual Detection and Object Recognition.....	20
Methodology	25
3.1 Defining Access Holes	25
3.2 Detecting Access Holes.....	27
3.2.1 Segmentation And Depth Interpolation.....	30
3.2.2 Attributes and Scoring.....	33

Depth Disparity.....	34
Hole Size.....	35
Photometric Brightness.....	36
3.2.3 Multi-Frame Score.....	37
Experiments and Results.....	41
4.1 Setup.....	41
4.2 Dataset.....	43
4.3 Evaluation.....	44
4.4 Discussion	48
Conclusion and Future Work.....	52
5.1 Summary	52
5.2 Limitations And Future Work.....	54
Bibliography	57

LIST OF TABLES

Table 4.1: AP using different numbers of segments.	45
Table 4.2: Individual attribute AP.	46
Table 4.3: Geometric and Photometric attributes' AP.	46
Table 4.4: Precision and AP with and without the use of multi-frame scoring.	47
Table 4.5: Number of possible holes for verification.	47

LIST OF FIGURES

Figure 2.1: Examples of collapsed reinforced and non-reinforced structures.	10
Figure 2.2: A dog and an USAR robot over rubble.	13
Figure 2.3: Microsoft Kinect active IR colour-depth sensor.	14
Figure 2.4: Stereolabs ZED dual camera 3D sensor.	15
Figure 2.5: Examples of ground robots.	16
Figure 2.6: Examples of sensor carrying UAVs.	17
Figure 2.7: Dog carried video systems.	19
Figure 2.8: A building collapse in Bangladesh with visible access holes.	22
Figure 3.1: Properties of an access hole.	26
Figure 3.2: The access hole detection pipeline.	28
Figure 3.3: Segmentation process example.	33
Figure 3.4: Sample Segmentations and output detections.	39
Figure 4.1: From left to right, P-R curves for initial segmentations of 3, 5 and 8.	45
Figure 4.2: Multi-frame matching P-R curves comparison.	47
Figure 4.3: An over segmented frame, resulting in false positives.	49

ABBREVIATIONS

3D	Three Dimensional
AP	Average Precision
CCD	Charge-Coupled Device
CNN	Convolutional Neural Network
CV	Computer Vision
DLP	Digital Light Processing
ERS	Entropy Rate Superpixel
FLIR	Forward Looking Infrared
GPU	Graphical Processing Unit
HUSAR	Heavy Urban Search And Rescue
IR	Infrared
P-R	Precision-Recall
POV	Point Of View
RGB	Red, Green, Blue
RRP	Reference Rubble Pile
SAR	Search And Rescue
UAV	Unmanned Aerial Vehicle
USAR	Urban Search And Rescue

CHAPTER 1

INTRODUCTION

Unfortunately, as a result of natural or man-made events, major emergency situations occur. Large emergencies that are beyond the ability of local emergency services to manage and require significant extra-jurisdictional planning and resources, are referred to as disasters. A disaster typically poses a threat to a large number of people in the area where it occurs. Depending on the nature of the disaster, it may pose multiple hazards to the population surrounding it. For example, the structural collapse of buildings poses danger to people both inside the structures and in close proximity to them. Collapsed buildings may crush people underneath the rubble as well as trap live victims inside gaps that form in the rubble. Under-rubble gaps are commonly referred to as “voids”. Voids are more likely to form in rubble as a result of structural collapses of reinforced structures, such as wood framed and steel reinforced concrete structures [1]; when reinforced structures collapse, large sections of flat surfaces such as walls and floors are likely to remain intact - hence when those flat objects collapse over other rubble, voids form underneath them. When a disaster strikes an urban area, entrapping victims under rubble, Urban Search And Rescue (USAR) teams are typically responsible for the ensuing response operations.

Following a disastrous event, emergency crews rush to the scene. Those responding teams are referred to as “first responders”. Emergency situations resulting from the collapse of

structures may cause victims to be entrapped in voids; in these scenarios a search for access points into the rubble and for signs of life of people trapped under the rubble must be completed before rescue operations can proceed. This search process is ideally performed by specially trained search specialists; these personnel must manually inspect the rubble to locate any areas of interest where victims may be trapped. In cases where USAR canine teams are available, they will be used to locate trapped and hidden live victims.

In some cases, the environment poses hazards which restrict access to first responders and canines [2]; in such situations, remotely controlled robotic technology, including ground robots and Unmanned Aerial Vehicles (UAVs), if available, may be used to continue the search; remotely controlled technology allows first responder teams to operate from a safe location [3]. Although robotic technology has been investigated for the purpose of USAR for many years, very few robots have been used in practice [3]. The first documented use of robots during an USAR operation was at the scene of the terrorist attacks on the World Trade Center in New York in 2001 [4].

Locating trapped victims and access points to under-rubble voids in a timely manner is vital for USAR operations. Over 90% of live victim extractions occur in the first 24 hours following a disaster, with a steep drop in the number of live extractions after the initial 48 hours following entrapment [1]. First responders must act promptly in order to find and extract live victims as quickly as possible. Access holes provide an entry point for first responders into rubble that may be near human-inhabited voids - creating a quick path to victims. However, not all victims may be reachable through access holes - in some cases, heavy rubble removal may be required in order to create an access path; the process of debris removal, particularly of heavy structural elements, requires the use of heavy machinery and especially trained Heavy Urban

Search And Rescue (HUSAR) personnel. Moreover, removal of rubble may endanger trapped victims by altering the stability of the fallen rubble, creating secondary collapses [1]. For these reasons, it is preferable to first remove any trapped victims quickly through available access holes without compromising rubble stability thus minimizing the risk to trapped victims and first responders.

A baseline for the automation of access hole detection in USAR disaster environments using a vision based technique was established by Kong et al. [5]–[7]. The system uses colour-depth sensors mounted on an UAV carrying agent that travels over the rubble; the colour-depth images are then processed by an algorithm which segments the images and marks regions of interest based on the properties of holes as extracted from the functional definition of an access hole’s utility. More specifically, Kong uses a Microsoft Kinect colour-depth sensor for the collection of data in his work. Both photometric information from the colour image as well as geometric information obtained from the depth component are used in order to detect holes. The detection of holes is based on an initial segmentation of the colour and depth components. Kong relies solely on depth information to perform segmentation; the segmentation obtained using the depth image is then applied to the colour image. Following segmentation, segments which are deemed sufficiently dark may be merged together, based on the assumption that holes generally appear darker to the colour sensor. Following segmentation and merging, each of the identified segments is then scored on five attributes identified as relevant characteristics of access holes: depth disparity, lack of brightness (darkness), brightness contrast, size and aspect ratio. Higher attribute scores equate to a higher confidence that a scored segment is an access holes.

Moreover, Kong claims that autonomously locating access holes helps alleviate some of the stressful work first responders are tasked with [7]; responders are able to use a single off-the-

shelf sensor in order to collect imagery of the rubble field, which is subsequently processed by the hole finding system to point out possible regions containing access holes which may lead to under-rubble voids and trapped victims.

1.1 OBJECTIVE

The work presented in this thesis aims to provide first responders with a complete and accurate system for the detection of access holes.

The work presented extends and improves on the initial research conducted by Kong [7]. The developed extension allows the system to process a continuous stream of frames produced by a sensor; the stream of frames aids the accurate recognition of access holes. Multiple frames providing multiple points of view of holes allow the system to make decisions based on multiple views of a region to provide better recognition. An improved segmentation and depth interpolation process, relying on the combination of photometric and geometric information, is also introduced; the combination of photometric and geometric components allows the system to be more resilient to corrupt and incomplete depth readings. These additions improve the accuracy of the hole finding algorithm in order to assist first responders, providing a fast and easy system for location and localization of access holes. The proposed system allows detected holes to be marked and tracked over a playback of the colour component of the dataset.

1.2 CONTRIBUTIONS

In order to help first responders perform their job more efficiently, the work of Kong [7] suggested the use of an automated hole finding algorithm. The work presented in this thesis improves upon Kong's algorithm by utilizing both colour and depth images during the segmentation process as well as relying on multiple frame to produce detection scores. A greatly improved accuracy has been achieved using the introduced enhancements.

My first contribution is the enhancement of the segmentation process by making use of both photometric and geometric information. In the previous work, segmentation of an input frame relied on the depth component alone. The depth component was over segmented using a superpixel segmentation algorithm followed by the joining of segments having an average brightness below a threshold. The work presented in this thesis makes use of both the colour and depth input components in order to segment the image into superpixels. Depth is interpolated within each of the segments using planes of best fit computed for each segment.

The previously developed autonomous access hole detection system provided a pipeline for the detection of access holes within single frames of the dataset. The frames were collected using a colour-depth sensor, which is capable of continuously recording frames, however only single frames were processed by the proposed system, limiting its practical use. The hole finding algorithm presented in this thesis accepts as input a stream of frames. My second contribution is an extension to the previous work where detected regions are matched using feature matching; this allows regions containing different views of a common access hole to be scored together and output as a single detection. By employing the multi-frame scoring mechanism, the number of

detections which require manual verification was reduced by over 73%. The number of output detections requiring manual verification by a human, likely a first responder, is greatly reduced.

My third contribution is an approach to the hole identification problem that offers improved accuracy and efficiency over the previous method. Using the presented enhancement and extension the system accuracy for perfect recall has been raised by 26%, while the number of output detections was reduced by over 73%.

1.3 ORGANIZATION

This thesis is organized into five chapters. Chapter 2, provides a review of previous research relating to the work presented in this thesis. Chapter 3 provides an in-depth discussion of the methodology and implementation of the autonomous access hole detection system and the proposed enhancements. In Chapter 4, multiple evaluations and assessments of the system are presented and discussed. Lastly, Chapter 5 presents a conclusion, summarizing the system's evaluation results, its abilities and limitations, followed by possible directions for future work.

CHAPTER 2

RELATED WORK

In this chapter, previous developments in USAR and Computer Vision (CV) relating to the work presented in this thesis are reviewed and discussed. First, an overview of USAR practices and findings is presented in Section 2.1. Following the domain introduction, the importance of voids within rubble is discussed in Section 2.1.1, highlighting the potential impact of the presented work; Section 2.1.2 presents background information for the detection of “signs of life” in the USAR domain. Acquisition of input data necessary for the autonomous access hole finding algorithm is discussed in Section 2.2. The hole finding system is not tied to a single type of sensor or sensing technology, hence potential sensor options are reviewed in Section 2.2.1, followed by a review of possible carriers for the sensor system, including ground robots, aerial robots and search dogs in Sections 2.2.2 through 2.2.4. Lastly, Section 2.3 explores visual object detection and recognition as well as methods to improve detection when dealing with imperfect sensor data.

2.1 URBAN SEARCH AND RESCUE (USAR)

Rubble environments caused by the structural collapse of buildings pose challenging circumstances for any form of mobile entity needing to access the environment, including humans, USAR dogs and ground and aerial robots; furthermore, such environments pose safety risks to people and dogs.

Depending on the availability of equipment and specialized teams at a disaster scene, an USAR search dog, optical and listening signs of life detection devices, ground robots or UAVs may be used to locate potential access holes and signs of life [5]–[11], ultimately leading to the rescue of trapped victims.

2.1.1 RUBBLE AND COLLAPSE SURVIVAL

A study of post-earthquake rescue operations indicated that after 48 hours have passed following a disaster, there is a significant decline in the number of victims rescued alive [1]. It is also important to note that an overwhelming 90 percent of live extrications occur within the first 24 hours after a collapse [1]; however, the study suggests there is a good chance of locating and rescuing live trapped victims in the first 5 days following a disaster. Macintyre et al. [1] also report that large voids are crucial to victim survival; large voids are more likely to occur in collapses of reinforced concrete or wood frame construction, where collapsed walls and floors tend to retain their shape and create voids. Large household appliances may also facilitate the

creation of under-rubble voids by supporting fallen rubble. When a rubble search area is vast, the use of USAR search and cadaver canines¹ is ideal [12]; no chemical sensor matches the scent detection abilities of a canine, moreover a canine can quickly move over a large area of rubble. Quickly locating regions likely to contain voids is crucial in order to concentrate available rescue personnel where victims may be entrapped [12]. Using information gathered from the literature and interviews of disaster managers, a framework for the prioritization of disaster search areas was developed [13].

Figure 2.1: Collapsed reinforced structure in Nairobi, Kenya (left) and collapsed non-reinforced brick buildings in Nepal following an earthquake (right). The scattered bricks of the non-reinforced structures in Nepal poorly facilitate the formation of large voids in the rubble they form.

¹ Cadaver canine teams are employed after search operations are complete during the ensuing recovery operation. The priority of search is to find live victims before any attempt is made to recover cadavers.

2.1.2 SIGNS OF LIFE

Voids within collapse rubble are important for the survival of trapped victims. However, locating voids and access holes in a large rubble field may be a strenuous task which may deprive first responders of precious time. In order to limit the area of search for access holes and voids, USAR teams prioritize areas of search based on multiple factors; the awareness of a live trapped victim is highly valued and plays an important role in the prioritization of search areas. In the context of USAR, signs of life refer to heartbeat, breathing and heat signatures, which may be detected using specialized equipment.

Microwave-based radar technology² has been used to search for vital signs of victims under rubble and behind walls [8], [11], [14]. These systems emit a microwave signal and receive back the reflected signal; using the Doppler effect, the radar systems are able to sense the breathing movements of a victim's chest [10]. The main disadvantage of such systems is their sensitivity to environmental noise from moving objects and rescuers [11]; some rubble materials, such as concrete, may also limit the range of detection for such radar systems.

Electronic acoustic listening devices pick up faint sounds of trapped victims [12]; these devices are used to survey large areas and by triangulation can pinpoint the location of a trapped victim. However, acoustic devices require a low level of ambient noise in order to be used. Once a suspected location of a trapped live victim is found, optical devices may be used to confirm the victim's condition and exact location [12].

² Often referred to as "Ground Penetrating Radar" or GPR.

Live human bodies generate heat; temperature differences between an environment surrounding a victim and a victim's body surface can be detected using thermal imaging devices such as Forward Looking Infrared (FLIR) cameras, however, other heat sources in the environment may contaminate the readings [12]. Infrared (IR) requires a direct line of sight, therefore thermal imaging is generally used to survey large open areas [15], [16]; an advantage of FLIR is its ability to see through dust, smoke and fog [12], [16][16], [20]. Thermal imaging systems used in USAR may have a pole mounted sensor [12] allowing the insertion of the sensor into rubble voids, with limited reach.

2.2 DATA ACQUISITION

The required data consists of a pair of colour and depth images; There are multiple sensors capable of collecting such data. In order to collect the data, a suitable sensor may be carried over rubble by various agents. Multiple robots have been designed for effective ground mobility and could be employed to traverse rubble to one extent or another; however, USAR environments sometimes prove too difficult for ground robots to traverse [4]. UAVs and USAR dogs may be more appropriate carrying agents for movement over rubble fields due to their mobility advantages and the dogs' sensory advantages [17]–[19].

Figure 2.2: A dog carrying equipment through rubble (left) and an USAR robot with a camera sensor over rubble (right).

This section reviews several possible methods for the collection of data from a rubble field. Section 2.2.1 provides a survey of sensors, including experimental and readily available colour-depth sensors that can provide the required input data pair of colour and depth images. The next three sections, 2.2.2 through 2.2.4, follow with discussions on the possible use of ground robots, UAVs and canines as sensor-carrying agents.

2.2.1 SENSORS

One of the most commonly used colour-depth devices employs a combination of calibrated colour and active IR sensors. Active IR sensors consist of an IR projector and camera in order to gauge depth. A popular and inexpensive example of this implementation is the Microsoft Kinect [20]. However, most applications of IR based colour-depth sensing are limited to indoor and low-light environments due to blinding of the IR camera by external IR sources such as the sun's rays [21].

Figure 2.3: Microsoft Kinect active IR colour-depth sensor

Another common sensor type, which is completely passive and has more robust behavior in IR flooded environments, is a stereo-camera. Stereo systems make use of two colour sensors which are typically set up using a fixed side-by-side placement. The resulting images are overlapped and a pixel disparity is computed [22]. However, because an algorithm is required to visually match points in one image to the other, stereo camera setups generally produce sparse or wrong depth information in areas consisting of certain textures, for example reflective surfaces, as well as objects in close proximity to the sensor [23]. The computed disparity between a stereo pair of images can then be used to estimate metric depth based on measured intrinsic and extrinsic camera geometry [24]. The cameras' intrinsic and extrinsic parameters must be carefully calibrated in order to allow for an accurate estimate of metric depth. Advancements in stereo vision allow for real-time, highly accurate Three Dimensional (3D) sensing using a hardware-based solution [25].

Figure 2.4: Stereolabs ZED dual camera 3D sensor

Another technique employing active sensing makes use of a Charge-Coupled Device (CCD) sensor combined with a structured light projection component, to allow for higher resolution depth and imagery [26]. This technique produces highly accurate data in chaotic, unstructured rubble environments using varying wavelengths of structured light.

2.2.2 GROUND-BASED ROBOTICS

Ground robots have been developed for use in USAR environments; however, not many have been used in practice due to the operational complexity they introduce, their limited applicability and their relative rarity. The first documented use of robots in an USAR operation was at the World Trade Centre disaster [4]. Ground robots have been proposed as sensors carrying agents for visual inspection and 3D mapping of interior voids in USAR disaster zones [26], allowing operators to visualize and investigate interior spaces within rubble as well as to locate victims.

Figure 2.5: Examples of ground robots. The AMOEBA-I shape-shifting and stair climbing robot (left, courtesy of [27]). Silver Bullet – a marsupial robot with its child robot deployed using a tether (middle, courtesy of [28]). A ground robot carrying a Digital Light Processing (DLP) projector and a CCD camera in an USAR environment (right, courtesy of [26]).

The operation of remote ground robots inside mines has aided in the detection of underground voids [29]; remote operation allows operators to stay within a safe environment while inspecting for underground spaces in mines. Similar approaches have been applied to USAR; navigation systems allowing ground robots to traverse difficult terrain within the USAR domain have been developed [30]. However, USAR rubble zones typically consist of chaotic terrain, limiting the extent to which ground robots are able to operate [31]. Marsupial and shape-shifting robots were developed as a more robust form of ground robot for traversal of and access to difficult rubble terrain [28]. However, the complexity of controlling shape-shifting robots makes them more difficult to employ in practice. Work continues to be done in this area including the proposal of an algorithm allowing a shape-shifting robot to climb stairs autonomously [27].

During operation many robots make use of a tether which transmits power and communications back and forth between the robot in rubble and the operator through a control unit; the tether may snag on rubble in USAR environments – therefore a tether freeing robot which travels along a tether, freeing it as it moves, was proposed [32]. However, the mobility abilities of ground robots remain inferior to those of humans and canines - limiting their operational adoption.

2.2.3 AERIAL ROBOTICS

Figure 2.6: A UAV carrying a sensor payload (left). Fixed wing UAV designed to carry an imaging sensor and telemetry equipment (right, courtesy of [33]).

UAVs have been used to perform visual inspection and analysis tasks of industrial equipment, which include small, enclosed spaces [34]. Approaches have been developed for the safe operation of UAVs for electrical line inspection as well as structural inspection in windy environments [35], [36]. A fixed-wing UAV designed to carry camera sensors can be used to provide a live view of affected areas below it, as well as provide a mosaic map using captured imagery [33]. Using UAVs for USAR operations allows a convenient mode of assessment that is

fast and safe [18]. The use of UAVs as sensor carrying agents has allowed modeling of rubble and collapsed buildings in 3D using a game engine, allowing first responders to virtually traverse and analyze a potentially hazardous environment from a safe location [37].

2.2.4 DISASTER DOGS

Dogs are amongst the best tools available to emergency response crews for the location of live victims (search and rescue dogs) and bodies (cadaver dogs) [9], [38]–[40]. USAR canine teams consist of a search and rescue (SAR) dog and a handler. Each handler is responsible to train their respective dog; dogs are trained to bark in a certain pattern to indicate a live victim is found. During SAR operations, a handler guides their dog through a search while the dog uses its sense of smell to locate hidden humans. During initial search operations, USAR canine teams sweep rubble in search of live survivors, while later searches involve cadaver dogs for the recovery of bodies.

Dog-carried video camera systems are used in various applications for policing and military operations, including post-disaster operations involving USAR. The mobility capabilities and agility of USAR dogs over difficult terrain, combined with the benefit of a powerful sense of smell, guides dogs to find trapped human victims very quickly. Therefore, canine teams remain the “gold standard” for USAR search operations [9]. The use of cameras on search dogs was first reported by [41], where a camera mounted to a dog collar transmitted wireless signals to a handheld display. A mechanical pan-and-tilt dog video system using two flank-mounted cameras has been demonstrated [17]. However, due to slow mechanical pan-and-

tilt speeds combined with signal loss and latency, the system was difficult to use in practice [42] and ultimately impractical given the demanding application. An improved Canine Augmentation Technology (CAT) system captures omni-directional video, allowing the operator a more complete view of a disaster scene [19]. The K9 Eye system, consisting of a head mounted sensor package, is currently commercially available for use in military, policing and SAR operations [43].

Figure 2.7: Dog carried video systems. A: Head mounted CAT system [44]. B: K9 Eye System [43]. C: CAT using dual mechanical pan & tilt camera domes [17]. D: View of rubble obtained using CAT 360 [19].

2.3 VISUAL DETECTION AND OBJECT RECOGNITION

In recent years there have been significant advancements in the field of CV with regards to object detection. Some of the most accurate detections were achieved using specially trained convolutional neural networks (CNN) [45]–[48], which are trained using many examples of the target object. CNNs appear to be a promising advancement toward the detection of objects that are hard to define clearly with regards to colour, shape, size and aspect ratio, such as holes. However, in order to train a CNN, a large number of varied training data is required, as well as a sufficient set of data for verification [49]; when the dataset available for training is not sufficiently large, the CNN will learn a representation which is over fitted to the examples provided during training. Within the problem domain of this thesis, it is difficult to acquire a sufficiently large and diverse data set. Hernández-López et al. [50] show that depth data can be used in combination with colour images to improve detection, at the same time allowing existing efficient methods to be used during segmentation of the colour images, producing a well performing algorithm. Processing colour and depth images independently at first, the algorithm produces labels based on the colour of a predefined target object. The label map is then overlaid with the depth map, in order to remove regions of similar colour that do not belong to the object of interest, making the assumption that object surfaces possess depth continuity. Another form of object recognition relies on shape matching; in this recognition method, a previously known shape is matched against unknown images to find matching shapes and corresponding objects [51]. Object detection based on a pre-defined model has been proposed by Brooks [52]; the target object models consist of a set of primitive shapes, such as cylinders, and a set of rules regarding the shapes' relative location. The model is capable of defining whole classes of objects

having slight variations in size and shape but composed of the same number of components in a similar alignment. A variation extending a form of model-based detection using parts proposed in [53] learns parts of a model and their variation in shape, size and location from training data marked using bounding boxes; this approach enables the detection system to include many object variations within a class, while dropping to need to explicitly define a model's parts.

Segmentation of input frames is often a first step in the process of object detection; segmentation is the process of dividing an image into multiple parts – segments. A basic segmentation technique is thresholding, where each pixel of an image is labelled as below or above the threshold. Multiple thresholds can be computed and used in order to label pixels with multiple classes [54]; this technique can be used to distinguish foreground objects from background. Superpixels are contiguous, homogeneous regions of an image; superpixel shapes and sizes tend to be irregular but are dependent on the algorithm used [55]. Superpixel segmentation can be formulated as a graph problem, where each pixel is a node; by performing a random walk on a graph and using a probabilistic model coupled with a balancing term, a superpixel segmentation with segments of similar sizes is achieved [55]. Tian et al. [56] used superpixel segmentations of MRI scans to form a 3D superpixel graph. The graph was then used to segment the 3D data and label segments. In order to accurately segment objects within an image with a matching depth map, [57] uses a colour image to over-segment a colour image and then relies on tangent planes fitted to the corresponding depth data in order to merge neighbouring segments which have smoothly connecting planes. Depth data gathered in uncontrolled environments may be sparse or corrupt, however, [57] and [58] show that high resolution colour imagery may be used in combination with sparse depth information in order to achieve a high accuracy object segmentation.

Figure 2.8: A building collapse in Bangladesh, 2013. Several access holes into the rubble can be seen.

Since a definition of shape, size and colour of access holes cannot be easily created to include all types of holes, the previously discussed methods are not a good fit for a hole finding system. However, it is fairly easy to functionally describe what access holes are:

All access holes have a sufficient depth, different from surrounding regions and have a certain minimum and maximum size and aspect ratios to allow a human to enter.

Objects of interest may be functionally defined to facilitate their detection [59]; instead of defining physical and visual traits of an object to be detected, a definition of an object's function and use is provided. The set of functionally defined properties are used as if-then rules to find objects matching the provided functional definitions. This approach facilitates detection of categories of objects which share function, but not necessarily the physical shape or visual appearance, such as holes. The work presented in this thesis is based on the demonstrated success of Kong et al. [7], where access holes are functionally defined using a set of visual cues and geometric constraints. Colour-depth imagery is used as input to the access hole finding algorithm; the imagery is segmented into superpixels before selected attributes are scored and

lumped into a final access hole likelihood score. The general process and attribute scoring is discussed in greater detail in Chapter 3.

CHAPTER 3

METHODOLOGY

This chapter presents a detailed breakdown of the technical approach employed by the access hole detection system. The chapter covers implementation details of new additions and enhancements as well as parts of the initial work by Kong which are used in the enhanced approach. Access holes are defined in the first section; the definition of an access hole is borrowed from the initial work by Kong [7]. The second section provides an overview of the access hole detection pipeline – followed by the rest of the chapter where technical details for each of the steps in the detection pipeline are presented.

3.1 DEFINING ACCESS HOLES

In order to aid first responders to quickly locate access holes, the developed algorithm relies on data collected using colour and depth sensors, which can be carried by any mobile agent. This allows rapid detection of holes that may be used to provide access to under-rubble voids.

A good description encompassing all types of holes of interest is needed in order to enable the detection of access holes. In order to define a hole for detection using a visual system,

a definition needs to include photometric properties, such as colour, texture and brightness, as well as geometric characteristics which include shape, size and depth. It is rather difficult to define all possible shapes, depths, colours and textures. Instead of providing a definition based on these attributes, an access hole is given the following functional definition based on its utility: An entry point which allows a searcher access to under-rubble voids.

Figure 3.1: An access hole, contoured with a green line. The hole is noticeably darker than the area surrounding it. The perpendicular red lines mark the width and length of the hole, used to calculate the hole's size and aspect ratio.

This thesis makes use of a functional access hole definition given by Kong [7]. An access hole in an USAR environment is defined as such a hole that allows access to an adult human searcher [7], with three attributes to define an access hole: “(I) depth disparity, (II) hole size and

(III) photometric brightness.” The depth, size and brightness properties of an access hole can be seen in Figure 3.1.

3.2 DETECTING ACCESS HOLES

In order to detect access holes, this thesis makes use of two key sequential data sets, colour imagery (video) and matching depth images. The detection of access holes is done by scoring segments within the image. Scores are provided for attribute extracted from the provided definition of an access hole – and scored are computed in the same manner as in [7]. For every given input frame, the algorithm segments the frame into segments using both the colour and depth components. Each of the resulting segments is given a score for each of the attributes identified. The individual attribute scores contribute to a final score given to each segment. Segments carrying a high combined score are considered to be potential access holes.

The input dataset is processed through several steps to achieve a final detection. The steps include (I) segmentation, (II) depth interpolation, (III) attribute scoring, (IV) matching segments across frames and (V) computing mean score for every segment. A diagram describing this flow is provided in Figure 3.2. The pipeline is similar in structure to that of the initial work, where the pipeline included steps (I) to (III), however, in the work presented by this thesis steps (I) and (II) are significantly changed and steps (IV) and (V) extend the original detection pipeline. Step (III) remains unchanged by the current work.

Figure 3.2: The detection pipeline. All input frames are segmented and scored individually first, this part of the process can be computed in parallel. Segment matching is then performed on all of the results of the individual frame detections before a final score is computed.

The first two steps of the pipeline, segmentation and depth interpolation, rely on a similar approach to that of the depth interpolation process described in [57]. In [57] an input frame’s colour image is over segmented, while depth information is used in order to join segments together to form larger segments; within each segment, a tangent plane is fitted to the depth data – neighbouring segments are joined together if their tangent planes connect smoothly. Following segment joining, depth is interpolated within each of the segments of the achieved final segmentation.

The first step in the detection pipeline is segmentation – where input colour and depth frames are segmented into superpixels. Ideally, every segment represents an entity, which can be an object, surface or hole. The segmentation step is designed to be performed using a segmentation technique which aims to produce segments that overlap with a single object only. In the work presented, segmentation is performed using Entropy Rate Superpixel segmentation [55], which was used by Kong as well, to allow a more direct comparison with the work done by Kong [7]. The colour and depth images are each initially segmented into N segments. First,

colour image segments are joined based on their brightness, while depth segments are joined based on their plane of best fit; this is followed by merging the segments of both colour and depth images. Section 3.2.1 presents a detailed discussion of the implementation used to produce the final segmentation.

The second step involves depth data interpolation within each segment - accommodating for missing and corrupt depth data. Based on an assumption, driven by domain information, that objects and surfaces are generally composed of flat regions, depth points are interpolated linearly. Depth interpolation is based on a flat plane fitted to the depth data within each of the segments achieved by the segmentation step. Linear interpolation is preferred for two main reasons: (I) Performance - flat surface fitting and interpolation is fast, and (II) Depth discontinuity - discontinuities in depth help to create object edges, which aid geometry based detection.

The third step in the detection pipeline is attribute scoring. In this step, each of the attributes determined as an important characteristic of access holes is given a score. More pronounced attributes dictate a higher score. The scores of all attributes are combined to form the final access hole score for every segment. The scoring process is described in detail in section 3.2.2.

Lastly, section 3.2.3 describes an extension of the initial work – multi-frame based scoring. For each of the potential holes detected, adjacent frames are searched for visually similar segments. Scores of segments across multiple frames containing the same hole are averaged to provide a more accurate score which takes into account multiple points of view. Holes with a high score are considered detected holes.

3.2.1 SEGMENTATION AND DEPTH INTERPOLATION

Before holes can be recognized, the input frames need to be segmented into segments corresponding to the different objects present in the scene. The algorithm treats the segments of the segmentation as distinct objects, performing hole recognition and identifying segments that may correspond to holes.

The method used to segment objects and interpolate depth across those segments is similar in nature to the work demonstrated in [57], where the colour image alone is used to define segments which are then joined based on the depth tangent planes of the individual segments. Based on the resultant segmentation, a new depth map is created by fitting a curved surface to the depth points within each of the segments, independently. In the implementation described below, both colour and depth images are segmented separately and segment joining is performed individually – using a separate method to merge colour and depth segments prior to overlaying the colour and depth segmentations.

The work presented performs two segmentations: the first uses the colour image only, while the second makes use of the geometric information found in the depth image. First, a segmentation is performed using only the colour image. The image is initially segmented into a number of segments – N . In order to join colour segments, a photometric based segment joining is performed, which is based on the darkness of segments, in a similar manner to the segment joining mechanism of the work presented by Kong. The average luminosity of a segment, extracted using the YUV colour space, is used as a gauge of brightness (or lack thereof) throughout this process. Neighbouring segments with a low average brightness level are joined

together, in an effort to create contiguous segments where holes, which are expected to be darker, exist.

Next, the depth image is segmented into N segments as well. Depth segments, which are based on the geometric makeup of the scene, are joined based on the similarity of their respective best fit planes. Once an initial segmentation is performed, the least squares method is used to fit a plane of best fit to the depth data points contained in each segment – the resultant plane closely resembles the surface of the object nestled in the segment. Adjacent segments having similar best fit planes are then joined together and a new plane is fitted to the now larger segment. To measure similarity between two segments, A and B, depth points for segment B are extracted from the best fit plane computed for segment A and compared with the actual depth points of segment B. A threshold is used to allow some room for error; because the size of the dataset is not sufficiently large to allow for a verification set, a threshold value was determined using a randomly selected set of five input frames. If points projected onto a plane fall close to the original depth points, within the threshold specified, the segments are joined. This process is done iteratively to ensure each segment's best fit plane is compared with all its neighbouring segments' planes.

Once both colour and depth images have been segmented and appropriate segments have been joined together, segments of both colour and depth images are overlaid to create a segmentation that is based on both photometric and geometric attributes. The segmentation overlay is computed by iterating over every pixel of the image. For each pixel, a segment label is computed based on the segment labels of the colour and depth images. Every pair of segments – one from the colour image and one from the depth image, imply a different segment label in the overlay segmentation.

The segments of the produced overlay segmentation are then joined based on the similarity of best fit planes fitted to each of the segments based on the depth data within the segment. This process is identical to the one employed for joining of the depth segmentation.

A simplified example of segmentation and overlay can be seen in Figure 3.3. The top left image shows the resulting segmentation of the colour image into 3 segments, while the top right image shows the resulting segmentation of the depth image into 3 segments. Since the image was segmented only into 3 segments, segment joining did not result in merging of any segments within the colour and depth segmentations. On the bottom row, the left image is the resulting segmentation of combining both the colour and depth segmentations. Planes of best fit were fitted to the resulting segmentation and based on similarity of planes of best fit, the final segmentation displayed in the bottom right image was obtained.

An assumption is made that USAR rubble of reinforced structures consists of mostly flat surfaces and objects; this is because structures built using steel reinforced concrete or wood frame tend to have flat walls and floors, which form voids when they collapse [12]. Based on this assumption, best fit planes are used to project depth data points throughout each of the segments identified, creating a new depth map to be used throughout the rest of the detection process; projecting planes allows a quick alternative to the interpolation of missing data points, and has the added benefit of smoothing outliers and incorrect depth readings.

Figure 3.3: Segmentation of colour image (top left), depth image (top right), overlaid segmentations (bottom left) and overlaid segmentations with merged segments (bottom right).

3.2.2 ATTRIBUTES AND SCORING

The presented work makes use of the attributes identified by Kong to characterize holes for the purpose of access hole detection. Attribute scores are computed using the same approach employed by Kong as well. This section provides a detailed overview of the key attributes identified and the methodology applied to compute their respective scores.

As previously discussed, three attributes were identified based on the functional utility definition of access holes: depth disparity, hole size and photometric brightness. These attributes serve as distinguishing factors for hole recognition. From these three attributes, a total of five score points were created: (i) Depth Disparity (ii) Hole Width (iii) Hole Aspect Ratio (iv) Brightness and (v) Relative Contrast. Each of the attribute scores describing an access hole are scored independently. In this section, an in-depth discussion of the attribute scoring process is provided. The input to each of the scoring functions is a pre-segmented colour-depth image, following the segmentation process described in Section 3.2.2. Every segment identified by the segmentation process is scored using the five metrics listed above. The goal is to recognize whether a given segment fits the provided definition of an access hole.

DEPTH DISPARITY

Depth differential is a key property in the definition of a hole. The depth disparity score S_d describes the difference between a segment's mean distance from the sensor and the mean of distances in its neighbouring segments. Minimum and maximum thresholds, D_{min} and D_{max} respectively, define the depth differential required for a segment to be considered a potential hole. The following formula defines the depth disparity score S_d for a depth disparity d :

$$S_d(d) = \begin{cases} 0, & d < D_{min} \\ 1, & d \geq D_{max} \\ \frac{d}{D_{max}}, & otherwise \end{cases}$$

If a segment's depth differential is lower than D_{min} , S_d is set to 0; if the depth differential is greater than D_{max} , S_d is set to 1; otherwise, S_d is the ratio of the segment's mean depth and D_{max} .

HOLE SIZE

The size of a hole is an important trait when searching for access holes. A hole too small may not allow rescuers or equipment entry into the rubble. Two scoring metrics are used to score a segment's size: the first is its width, and the second is its aspect ratio. The width, simply defined as the longest straight path across a segment, is measured for each segment. The width score S_w is based on a pre-set threshold for the minimum required width for a hole; if a segment's width is larger than the threshold – the segment receives a score of 1, if the width is smaller than the threshold – the score is a ratio of the width and the width threshold.

The second hole size score, the aspect ratio score S_r , is closely related to the width score. The aspect ratio is calculated by dividing the total area of the segment R_A , by $R_W * R_L$, where R_W is the segment's width and R_L is the segment's length.

$$S_r = \frac{R_A}{R_W * R_L}$$

The above formula computes the ratio of segment area versus a rectangular segment with the same width and length – accounting for irregularly shaped segments. Higher aspect ratio scores indicate a more coherently shaped segment that is more likely to enable access.

PHOTOMETRIC BRIGHTNESS

Photometric brightness was identified as a key visual cue of not only access holes, but of holes in general. The assumption is that deep holes generally have less exterior light entering into them, and as a result appear darker. Two metrics are used to gauge photometric brightness. The first is the absolute brightness of a segment. The second is a score of a segment's brightness relative to its neighbouring segments – a segment's contrast.

In order to calculate brightness based scores, the colour image is converted from the Red, Green, Blue (RGB) colour space into the YUV colour space. The YUV colour space is composed of three elements: a 'luma' component Y corresponding to luminosity, and two colour components U and V , referred to as the blue-difference and red-difference, respectively. For each segment, two independent scores are calculated in order to gauge photometric brightness; the first score is a score of darkness, or absolute brightness, while the second is a score of contrast relative to neighbouring segments. For the calculation of both scores, the luminosity component of the colour image is used as an indicator of brightness.

To calculate a segment's absolute brightness score, S_b , an upper limit is set for the maximum brightness value; the average brightness is then divided by the upper limit, presenting a ratio of the segment's brightness. A higher ratio represents brightness, although darkness is of interest, hence the calculated brightness score is subtracted from 1 to provide the segment's darkness score.

Next, a score is determined for relative contrast, S_c , based on the average brightness differential between a segment and its surrounding neighbours. To calculate the brightness

differential, first the average brightness intensities of all neighbouring segments are averaged; then, the average brightness intensity of the segment of interest is deducted from it. The calculation for the score of relative contrast is performed in a similar manner to the absolute brightness score calculation. The segment's brightness differential is divided by a predetermined value which corresponds to the brightness differential required for a segment to qualify as a possible hole; the result of this division provides the relative contrast score, which is essentially a ratio of the segment's brightness differential versus the ideal brightness differential.

The two scores (brightness score and contrast score) make up equal parts of the photometric brightness score.

3.2.3 MULTI-FRAME SCORE

This thesis demonstrates the use of a naive multi-frame scoring system in order to improve recall and accuracy. The hole detection algorithm performs a single pass over all data frames involving segmentation, depth interpolation and attribute scoring. The algorithm identifies all segments that may be holes. Another pass is then performed over the output detections – matching segments containing the same hole across multiple frames in order to produce the final score for each hole detected.

As the exact shape of a hole is unknown, only un-occluded parts of holes visible to the camera are available for the algorithm to base its scoring. Holes within each frame are scored by relying on visible information at that frame. Scores for a segment are collected from multiple

frames where the segment is visible and averaged to create a more robust score. This new score allows for more viewpoints to be considered, potentially eliminating obstructions. Using multiple frames can potentially help combat photometric nuances including blur and glare, as well as corrupted depth data.

Once potential holes have been recognized within all processed frames, segments containing those potential holes are matched with segments containing the same hole in other frames within a predefined neighbourhood. A smaller neighbourhood will allow the algorithm to conclude sooner, while risking loss of potential frame matches. Extracted features from segments are used by the algorithm to match segments across multiple frames. Many feature extraction algorithms may be suitable for use in the extraction step, some popular options include SIFT, SURF and ORB [60]–[62]; the current Matlab implementation extracts minimum Eigenvalue features [63] and uses Fast Retina Keypoints (FREAK) [64] descriptors to perform feature matching. Segments containing a sufficient number of matching features are considered to contain a view of the same hole. Scores of all frames containing a view of the same hole are then averaged, creating the final hole score.

Figure 3.4: Examples of segmentations and output detections. Top row shows examples of segmentations where the input colour and depth frames were segmented into eight segments. The bottom row displays respective output detections using bounding boxes over the colour images.

CHAPTER 4

EXPERIMENTS AND RESULTS

In this chapter, a variety of experiments are defined, each showcasing an aspect of the system, and their results are examined. The system has been tested and verified in various ways. The work presented in this thesis was tested and compared to the single frame detection functionality of the hole finding system introduced by Kong [7]; the results highlight the importance of the segmentation process introduced. The multi-frame enhancement was also tested and compared with both single-frame performance and Kong's results. All tests were performed using the dataset introduced by Kong to allow a fair basis for comparison.

4.1 SETUP

This thesis introduces an improved segmentation process which relied on colour and depth information. First, segmentation and segment joining are performed on both colour and depth frames independently. Within the colour image, the segment joining process makes use of the same approach used by Kong in his work – neighbouring segments were joined together if their average brightness was below a threshold; the threshold value was set to the same value used in Kong's work, a luma value of 0.274. Depth frame segment joining was based on

similarity of best fit planes, where two neighbouring segments were joined if the average distance of their depth points from the neighbouring segment's plane was below a threshold. This threshold was determined based on observed sensor accuracy and error. Since the dataset is of a limited size, it was not possible to divide the data into training and verification sets in a meaningful manner. Therefore, the threshold value was empirically selected using a randomly selected set of five frames. The same depth threshold is also used to join segments within the segmentation obtained by overlaying both colour and depth segmentations.

To allow for an accurate comparison with results of the previous work reported by Kong [7], parameter borrowed from the initial work set to the same values as in Kong's work. The minimum and maximum depth thresholds for the depth score, S_d , were set to 200mm and 1951mm; hole width and length thresholds, used for the calculation of S_w and S_r , were set to 655mm and 368mm. These values for the depth, width and length parameters were based on the average measurements of an adult human, as reported in [65]. Further, the brightness threshold used for the computation of the brightness and relative contrast scores S_b and S_c was assigned a constant value of 0.274; the brightness threshold was learned by Kong from multiple images of holes in various scenarios from a range of different cameras [7].

Segment scores are made up of the five individual attribute scores S_d, S_w, S_r, S_b and S_c using the following formula:

$$S = (W_d S_d + B_d) + (W_w S_w + B_w) + (W_r S_r + B_r) + (W_b S_b + B_b) + (W_c S_c + B_c)$$

Each of the weights W_i is given an equal weight of $1/5$ and a bias B_i of 0. The segment score is effectively the average of individual attribute scores. The work presented by Kong was verified using a similar setup, allowing a level playing field for the comparison of the work

presented with the system proposed by Kong. In order to evaluate the performance and accuracy of the work proposed in this thesis, multiple metrics were recorded.

4.2 DATASET

The work presented in this thesis extends and enhances the initial work presented by Kong [7]. To allow for a meaningful comparison between the two works, evaluations presented in this chapter make use of the dataset introduced through Kong's work. This dataset was recorded at the Reference Rubble Pile (RRP) of the Ontario Provincial Police in Bolton, Ontario. The dataset was collected using an ASUS Xtion Pro colour-depth camera mounted underneath a UAV. The Asus Xtion Pro produces colour-depth images at a resolution of 640x480; the colour images have 32 bit colour depth, and the produced depth images provide per-pixel depth. The dataset includes a total of 254 colour-depth images, of which 166 images cover multiple views of 18 unique holes [7].

An attempt was made to collect additional data from rubble environments. The RRP of the OPP where Kong previously collected his data has been declared structurally unsafe, as such, the collection of data could only be accomplished by walking around the boundaries of the RRP. Two sensors were employed to collect a small dataset, a Microsoft Kinect (active IR) and a Stereolabs ZED (stereo camera). Since it was not possible to get close to the RRP, the resolution and quality of the footage captured using the Kinect sensor (active IR) was inadequately low and insufficient for the algorithm to function as designed. Similarly, the depth map computed by the ZED camera was too sparse to distinguish any holes; the raw stereo footage may be usable with

further, more sophisticated processing to produce a more dense depth map, which will allow the data to be used; however, this is outside the scope of the work and was not attempted.

4.3 EVALUATION

This section presents multiple evaluations showcasing the performance and accuracy of the proposed work. First, a direct comparison with Kong’s initial work is presented; the comparison is based solely on individual frames’ segment hole scores, without the use of multi-frame enhancement. The algorithm’s performance is measured using Precision-Recall (P-R) [66]. Let TP be the total number of true positives – the number of detections which correctly identify holes. Let FP be the total number of false positives – the number of detections which incorrectly identify holes. Then, the precision P is defined by the following formula:

$$P = \frac{TP}{TP + FP}$$

and the recall, R is defined by

$$R = TP/nP$$

where nP is the total number of holes present in the dataset. Ground truths for access holes in the dataset were hand labeled by Kong using rectangular bounding boxes. Detections are considered to be correct if the detection bounding box overlaps with at least 50% of the area in a ground truth bounding box. Table 4.1 shows a measure of Average Precision (AP) based on the number of initial segments used. The current implementation uses Entropy Rate Superpixel (ERS) [55]

segmentation to segment both colour and depth frames. The same superpixel segmentation algorithm was used throughout Kong’s work.

Number of Segments	3	4	5	6	7	8	9	10
Average Precision	0.65	0.44	0.59	0.50	0.44	0.47	0.33	0.35

Table 4.1: AP using different numbers of segments during initial segmentation of the colour and depth images.

Figure 4.1: From left to right, P-R curves for initial segmentations of 3, 5 and 8.

Although it may seem that a low number of segments should be preferable based on the information in Table 4.1, segmentations using less than 7 segments failed to achieve a 100% recall. Based on this information, all images were segmented into 8 segments throughout the rest of the presented analysis. The resulting P-R curves for an initial segmentation of 3, 5 and 8 segments are displayed in Figure 4.1. Table 4.2 shows the contributions of individual attribute scores to the AP, while Table 4.3 shows photometric, geometric and combined APs.

	Depth	Width	Aspect Ratio	Brightness	Relative Contrast
Average Precision	0.24	0.26	0.08	0.36	0.35

Table 4.2: Individual attribute AP.

	Geometric	Photometric	Combined
Average Precision	0.37	0.45	0.47

Table 4.3: Geometric and Photometric attributes' AP.

Multi-frame scoring was introduced in order to further reduce the number of false positives. The aim is to limit the number of detections for holes present in the dataset, ideally reporting each true hole only once, rather than once per frame of appearance. In order to test the performance of the multi-frame scoring technique, the criteria for Recall was slightly modified; the requirement is that every access hole in the dataset is marked at least once. In Figure 4.2, the P-R curve with and without multi-frame matching is displayed. Table 4.4 displays the precision achieved for a 100% recall rate as well as the AP with and without the use of multi-frame scoring, with segmentations of 5 and 8. When multi-frame scoring is not used, each frame is scored individually; every segment which has a confidence score higher than 0 is output as a possible detection, leading to a large number of potential holes. Multi-frame scoring scores segments across multiple frames where matching segments are found; by only outputting a single score for potential matched segments, the number of output detections is significantly reduced.

Table 4.5 shows the effects of multi-frame scoring on the number of output detections – the numbers represent the total number of detections for all non-zero confidence levels.

Figure 4.2: P-R curves with multi-frame matching (left) and without multi-frame matching (right).

Number of Segments	5		8	
Multi-frame scoring	Yes	No	Yes	No
Perfect recall Precision	0.42	-	0.31	0.25
Average Precision	0.62	0.59	0.54	0.47

Table 4.4: Precision and AP with and without the use of multi-frame scoring.

Number of Segments	5		8	
Multi-Frame Scoring	Yes	No	Yes	No
Number of segments to verify	758	2464	1102	4107

Table 4.5: Number of possible holes for verification.

4.4 DISCUSSION

The main goal of the work presented in this thesis is to improve upon the initial work laid out in [7] and to provide a more accurate tool for first responders. A notable improvement has been achieved, as the highest AP reported by Kong is 0.43, while the work presented in this thesis achieves an AP of 0.65 as demonstrated in Table 4.1 – an improvement of over 22% and a 51% more precise result; in fact, the only reported APs which are lower than the highest AP of the previous work reported in [7] are 0.33 and 0.35 for segmentations using 9 and 10 segments, which over-segment images into small segments, producing more false positives and false negatives. An example of an over segmented frame can be seen in Figure 4.3. Interestingly, a higher AP is achieved with a lower number of segments, both in the work presented here and in the results reported by Kong.

The highest AP was obtained with the number of segments for initial segmentation set at 3. As the total number of detections per frame is bound by the number of segments, a smaller number of segments in the final segmentation helps curb the number of possible false positives. Requiring the segmentation algorithm to produce such a small number of output segments also forces more significant differences between those segments; therefore, deep holes and very distinctive objects tend to stand out in the segmentation. However, with a small number of segments, some holes may be missed, as the segments may be too broad; when segmenting into less than 7 segments, the algorithm was unable to achieve complete recall, missing some holes. For example, when the segmentation was set to 5 segments, the algorithm missed 3.33% of holes. A false negative ratio of 3.33% may be acceptable in certain domains; however, in the domain of this work, it may mean a potential loss of life – hence ideally a higher number of

segments should be used in order to avoid missed detections. Using a segmentation into 8 segments, the precision for a perfect recall greatly increases from the reported results of Kong – from 0.16 to 0.25.

Figure 4.3: An over segmented frame, resulting in false positives and false negatives. The colour and depth components were segmented into ten segments.

All of the individual attribute scores' APs, with the exception of aspect ratio, outperform the corresponding attribute APs in Kong's work. Geometric attribute scores do not perform very well on their own, producing APs ranging from 0.08 to 0.26. However, when the three attribute scores are combined, the geometric attribute scores alone manage to reach an AP as high as 0.37. On the other hand, the photometric attribute scores individually have APs of 0.35 and 0.36 for relative contrast and brightness, respectively; when the two attribute scores are combined,

photometric attribute scores manage a respectable AP of 0.45. Together, all five geometric and photometric attribute scores (discussed in Section 3.3) reach an even higher combined AP of 0.47.

The AP curve resulting from the use of multi-frame scoring is shown in Figure 4.2. The AP score using multi-frame scoring with an initial segmentation into 8 segments rises from 0.47 to 0.54 – a significant improvement. The multi-frame scored segments manage to also considerably raise the precision for a perfect recall; with a 100% recall, the precision increased from 0.25 to 0.31. It is also important to note that the number of potential holes identified for consideration substantially reduced from 4107 to 1102, cutting the number of segments for consideration by over 73%.

Using initial segmentations having less than 7 segments failed to recall all holes on a frame-by-frame basis; however, using a segmentation into 5 segments with multi-frame scoring, all access holes found in the dataset were accounted for. The results can be seen in Table 4.4 and Table 4.5. The AP was measured at 0.62 while precision for perfect recall was significantly increased to 0.42. The number of output detections was considerably reduced from the 1102 produced using a segmentation into 8 segments to 758.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 SUMMARY

This thesis expands initial work on access hole detection toward a fully autonomous access hole finding algorithm. An autonomous access hole detection system would greatly benefit first responders in their search for survivors trapped under rubble. Implemented improvements to the original work by Kong [7] aim to reduce false positives as well as false negatives, to provide a more meaningful set of output detections for operator verification. An improved segmentation process tailored to photometric and geometric attribute scoring is introduced. Scoring of segments based on multiple frame views of a region was added as well, allowing a far more accurate score based on several points of view.

A major part of the enhancements to the hole finding system is related to segmentation and merging of segments, in preparation for detection. The newly introduced segmentation process segments colour and depth images individually and overlays the resulting segmentations. The current implementation uses ERS [55] to segment both the colour and depth images. The new segmentation minimizes segmentations where a single segment overlaps a hole and its

surroundings, leading to more accurate attribute scores. The results showcased in Chapter 4 highlight the importance of a segmentation and segment merging process that is tailored to the domain. Individual segments of the final segmentation are candidates for detection, therefore each segment produced by segmentation should ideally be comprised of a single object; however, most importantly, segments should closely overlay access holes, and a segment should not overlay both a hole and surfaces adjacent to the hole, in order to avoid false negatives. Hence, segmentation should be treated as a core component of the access hole detection system. The highest AP achieved was 0.65 compared to 0.43 previously through Kong's work. The highest AP using a segmentation achieving perfect recall was 0.47. Precision for complete recall was raised from 0.16 to 0.25 – a considerable improvement.

Scores based on multiple frames increase the utility of the system as a whole by limiting the number of output detections; furthermore, multi-frame scoring achieves similar or higher AP and complete recall precision. Multi-frame scores are based on segment matching performed by extraction of features from all segments and matching with features of segments in other frames. The current implementation uses minimum Eigenvalue features [63]. As a result of multi-frame matching, where the goal is to recall each available hole at least once, the number of output detections was reduced from 4107 to 1102 when a segmentation of 8 was used. When a segmentation of 5 was used, the number of output detections was further reduced to 758. Although segmentation into 5 initial segments did not facilitate a complete recall of all access holes in the dataset, multi-frame scoring managed to mark every hole at least once; thereby, an AP of 0.62 and a perfect recall precision of 0.42 were achieved.

The higher accuracy achieved, together with the reduction in the number of output detections should allow for a much quicker and easier manual verification process for detections.

5.2 LIMITATIONS AND FUTURE WORK

The hole finding algorithm presented in this thesis aims to locate access holes in a timely manner, allowing first responders to tend to other tasks on the field. The system receives as input a set of input colour and depth frames gathered using a carrying agent over rubble. The data is off-loaded onto a computer where it is processed. While the enhancements presented substantially improve the overall accuracy, processing time was not addressed. The system requires multiple seconds per frame for processing. Therefore, the use of parallelization and Graphical Processing Units (GPU) which are expected to improve the performance considerably and should be considered for future implementations.

In its current format, the system is not able to be run in real-time using streaming data. The multi-frame scoring component needs to be reconsidered and implemented in a way that will allow real-time detection. Perhaps a moving window approach can be used to consider multi-frame scores over a neighbourhood of several frames; the most recent frame will check for matching segments over only the last recorded N frames and produce a score based on those matches.

The results achieved by the presented work relied on the introduction of a better suited segmentation algorithm as well as a means of consolidating hole scores across multiple frames. However, attributes and their scoring methods were kept consistent with the original work by Kong [7]. To further improve accuracy, the system would benefit from the addition of attributes. Weighted sums of attribute scores may also benefit the system; however, there is currently insufficient data to learn the weights of attributes.

Detections made by the system are displayed frame by frame, or may be overlaid atop a playback of the video used during the process. However, to promote a more natural and easier localization of access holes in the field, localization and mapping of potential holes over a 3D model of the rubble should be considered.

BIBLIOGRAPHY

- [1] A. G. Macintyre, J. A. Barbera, and E. R. Smith, "Surviving Collapsed Structure Entrapment after Earthquakes: A 'Time-to-Rescue' Analysis," *Prehospital and Disaster Medicine*, Feb-2006. .
- [2] S. M. Gwaltney-Brant, L. A. Murphy, T. A. Wismer, and J. C. Albretsen, "General toxicologic hazards and risks for search-and-rescue dogs responding to urban disasters," *J. Am. Vet. Med. Assoc.*, vol. 222, no. 3, pp. 292–295, Feb. 2003.
- [3] Z. Wang and H. Gu, "A review of locomotion mechanisms of urban search and rescue robot," *Ind. Robot Int. J.*, vol. 34, no. 5, pp. 400–411, 2007.
- [4] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 33, no. 3, pp. 367–385, 2003.
- [5] C. Kong, A. Ferworn, J. Tran, S. Herman, E. Coleshill, and K. G. Derpanis, "Toward the automatic detection of access holes in disaster rubble," in *2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2013, pp. 1–6.
- [6] C. Kong, "Discovering access holes in disaster rubble with functional and photometric attributes," Ryerson University, 2015.
- [7] C. Kong, A. Ferworn, E. Coleshill, J. Tran, and K. G. Derpanis, "What is a Hole? Discovering Access Holes in Disaster Rubble with Functional and Photometric Attributes: What is a Hole? Discovering Access Holes in Disaster Rubble with Functional and Photometric Attributes," *J. Field Robot.*, vol. 33, no. 6, pp. 825–836, Sep. 2016.
- [8] M. Donelli, "A rescue radar system for the detection of victims trapped under rubble based on the independent component analysis algorithm," *Prog. Electromagn. Res. M*, vol. 19, pp. 173–181, 2011.
- [9] A. Agapiou, P. Mochalski, A. Schmid, and A. Amann, "Potential Applications of Volatile Organic Compounds in Safety and Security," in *Volatile Biomarkers*, Elsevier, 2013, pp. 514–558.
- [10] I. Y. Immoreev, S. Samkov, and T.-H. Tao, "Short - distance ultrawideband radars," *Aerosp. Electron. Syst. Mag. IEEE*, vol. 20, no. 3, pp. 9–14, 2005.
- [11] I. Arai, "Survivor search radar system for persons trapped under earthquake rubble," in *APMC 2001. 2001 Asia-Pacific Microwave Conference (Cat. No.01TH8577)*, 2001, vol. 2, pp. 663–668 vol.2.
- [12] M. Statheropoulos *et al.*, "Factors that affect rescue time in urban search and rescue (USAR) operations," *Nat. Hazards*, vol. 75, no. 1, pp. 57–69, 2015.
- [13] R. Hassanzadeh and Z. Nedovic-Budic, "Where to go first: prioritization of damaged areas for allocation of Urban Search and Rescue (USAR) operations (PI-USAR model)," *Geomat. Nat. Hazards Risk*, vol. 7, no. 4, pp. 1337–1366, 2016.
- [14] D. Dei *et al.*, "Non-Contact Detection of Breathing Using a Microwave Sensor," *Sensors*, vol. 9, no. 4, pp. 2574–2585, Apr. 2009.
- [15] H. K. Jusoff, "Pixel-based airborne hyperspectral sensing technique for search-and-rescue of the missing RMAF NURI helicopter in Genting-Sempah, Malaysia," *Disaster Prev. Manag.*, vol. 19, no. 1, pp. 87–101, 2010.

- [16] B. Wang, L. Dong, M. Zhao, H. Wu, Y. Ji, and W. Xu, "An infrared maritime target detection algorithm applicable to heavy sea fog," *Infrared Phys. Technol.*, vol. 71, pp. 56–62, Jul. 2015.
- [17] J. Tran, A. Ferworn, C. Ribeiro, and M. Denko, "Enhancing canine disaster search," in *2008 IEEE International Conference on System of Systems Engineering*, 2008, pp. 1–5.
- [18] M. Onosato, F. Takemura, K. Nonami, K. Kawabata, K. Miura, and H. Nakanishi, "Aerial robots for quick information gathering in USAR," in *2006 SICE-ICASE International Joint Conference*, 2006, pp. 3435–3438.
- [19] B. Waismark, A. Ferworn, and M. Scanlan, "CAT 360 - Canine augmented technology 360-degree video system," in *2015 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2015, pp. 1–4.
- [20] Zhengyou Zhang, "Microsoft Kinect Sensor and Its Effect," *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, 2012.
- [21] J. Hernandez-Aceituno, R. Arnay, J. Toledo, and L. Acosta, "Using Kinect on an Autonomous Vehicle for Outdoors Obstacle Detection," *IEEE Sens. J.*, vol. 16, no. 10, pp. 3603–3610, 2016.
- [22] F. Gonzalez and R. Perez, "Neural mechanisms underlying stereoscopic vision," *Prog. Neurobiol.*, vol. 55, no. 3, pp. 191–224, 1998.
- [23] A. Broggi *et al.*, "The passive sensing suite of the TerraMax autonomous vehicle," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 769–774.
- [24] Z. Zhang, "Determining the Epipolar Geometry and its Uncertainty: A Review," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 161–195, 1998.
- [25] K. Ambrosch and W. Kubinger, "Accurate hardware-based stereo vision," *Comput. Vis. Image Underst.*, vol. 114, no. 11, pp. 1303–1316, Nov. 2010.
- [26] B. Mobedi and G. Nejat, "3-D Active Sensing in Time-Critical Urban Search and Rescue Missions," *IEEEASME Trans. Mechatron.*, vol. 17, no. 6, pp. 1111–1119, Dec. 2012.
- [27] B. Li, J. Chang, and L. Zhu, "Research on autonomous stairs climbing for the shape-shifting robot," in *2015 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2015, pp. 1–6.
- [28] R. R. Murphy, "Marsupial and shape-shifting robots for urban search and rescue," *IEEE Intell. Syst. Their Appl.*, vol. 15, no. 2, pp. 14–19, 2000.
- [29] S. S. Wilson, L. Gurung, E. A. Paaso, and J. Wallace, "Creation of robot for subsurface void detection," in *2009 IEEE Conference on Technologies for Homeland Security*, 2009, pp. 669–676.
- [30] A. Sinha and P. Papadakis, "Mind the gap: detection and traversability analysis of terrain gaps using LIDAR for safe robot navigation," *Robotica*, vol. 31, no. 7, pp. 1085–1101, Oct. 2013.
- [31] A. Ollero and L. Merino, "Control and perception techniques for aerial robotics," *Annu. Rev. Control*, vol. 28, no. 2, pp. 167–178, Jan. 2004.
- [32] A. Specian and M. Yim, "Friction binding study and remedy design for tethered search and rescue robots," in *2015 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2015, pp. 1–6.
- [33] G. Nugroho *et al.*, "Development of a Fixed Wing Unmanned Aerial Vehicle (UAV) for Disaster Area Monitoring and Mapping, Development of a Fixed Wing Unmanned Aerial Vehicle (UAV) for Disaster Area Monitoring and Mapping," *J. Mechatron. Electr. Power*

- Veh. Technol. J. Mechatron. Electr. Power Veh. Technol.*, vol. 6, no. 2, pp. 83–88, Dec. 2015.
- [34] J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Huerzeler, and R. Siegwart, “A UAV system for inspection of industrial facilities,” in *2013 IEEE Aerospace Conference*, 2013, pp. 1–8.
 - [35] C. Liu, Y. Liu, H. Wu, and R. Dong, “A Safe Flight Approach of the UAV in the Electrical Line Inspection,” *Int. J. Emerg. Electr. Power Syst.*, vol. 16, no. 5, pp. 503–515, 2015.
 - [36] J. Guerrero and Y. Bestaoui, “UAV Path Planning for Structure Inspection in Windy Environments,” *J. Intell. Robot. Syst.*, vol. 69, no. 1–4, pp. 297–311, 2013.
 - [37] S. Herman, “Disaster Scene Reconstruction: Modeling, Simulating, And Planning In An Urban Disaster Environment,” 2014.
 - [38] R. Douvillier, “Teamwork in Action,” *Firehouse*, vol. 36, no. 6, pp. 92–94, Jun-2011.
 - [39] P. Snodgrass, “The Fog of Disaster: Into Haiti with Two Tampa Fire Usar Members,” *Firehouse*, vol. 35, no. 3, pp. A6–A9, Mar-2010.
 - [40] L. Collins, “USAR Response to Japan Earthquake and Tsunamis, Part 2,” *Fire Engineering*, vol. 164, no. 11, pp. 75–80, Nov-2011.
 - [41] J. German, “News Release - K-9 camera.” [Online]. Available: <https://share.sandia.gov/news/resources/releases/2002/K9cam.htm>. [Accessed: 18-Dec-2016].
 - [42] M. Gerdzhev, J. Tran, A. Ferworn, K. Barnum, and M. Dolderman, “A scrubbing technique for the automatic detection of victims in urban search and rescue video,” in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, 2010, pp. 779–783.
 - [43] “Military & Law Enforcement Division - K9-eye Dog Camera System - Video Technology.” [Online]. Available: <http://www.videotechnology.tv/military-law-enforcement-division/k9-eye-dog-camera-system>. [Accessed: 02-Jan-2017].
 - [44] A. Ferworn *et al.*, “Urban search and rescue with canine augmentation technology,” in *2006 IEEE/SMC International Conference on System of Systems Engineering*, 2006, p. 5 pp.-pp.
 - [45] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *ArXiv150601497 Cs*, Jun. 2015.
 - [46] K. Kang *et al.*, “T-CNN: Tubelets with Convolutional Neural Networks for Object Detection from Videos,” *ArXiv160402532 Cs*, Apr. 2016.
 - [47] G. Cheng, P. Zhou, and J. Han, “Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
 - [48] Y. Zhu, R. Urtasun, R. Salakhutdinov, and S. Fidler, “segDeepM: Exploiting segmentation and context in deep neural networks for object detection,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4703–4711.
 - [49] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,” *Neurocomputing*.
 - [50] J.-J. Hernández-López, A.-L. Quintanilla-Olvera, J.-L. López-Ramírez, F.-J. Rangel-Butanda, M.-A. Ibarra-Manzano, and D.-L. Almanza-Ojeda, “Detecting objects using color and depth segmentation with Kinect sensor,” *Procedia Technol.*, vol. 3, no. Complete, pp. 196–204, 2012.
 - [51] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 24, no. 4, pp. 509–522, 2002.

- [52] R. A. Brooks, "Model-Based Three-Dimensional Interpretations of Two-Dimensional Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 2, pp. 140–150, 1983.
- [53] J. M. Gonfaus, M. Pedersoli, J. González, A. Vedaldi, and F. X. Roca, "Factorized appearances for object detection," *Comput. Vis. Image Underst.*, vol. 138, pp. 92–101, Sep. 2015.
- [54] V. Osuna-Enciso, E. Cuevas, and H. Sossa, "A comparison of nature inspired algorithms for multi-threshold image segmentation," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1213–1219, Mar. 2013.
- [55] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104.
- [56] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "Superpixel-Based Segmentation for 3D Prostate MR Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 3, pp. 791–801, Mar. 2016.
- [57] K. Matsuo and Y. Aoki, "Depth Interpolation Using Tangent Planes on Superpixels of a Color Image," *Electron. Commun. Jpn.*, vol. 98, no. 12, pp. 17–29, Dec. 2015.
- [58] B. Dellen, G. Alenyà, S. Foix, and C. Torras, "Segmenting color images into surface patches by exploiting sparse depth data," in *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, 2011, pp. 591–598.
- [59] P. H. Winston, T. O. Binford, B. Katz, and M. Lowry, *Learning physical descriptions from functional definitions, examples, and precedents*. Department of Computer Science, Stanford University, 1983.
- [60] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [61] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [62] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International conference on computer vision*, 2011, pp. 2564–2571.
- [63] J. Shi and C. Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.
- [64] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.
- [65] J. Panero and M. Zelnik, *Human dimensions & interior space: a source book of design reference standards*. New York, N.Y: Whitney Library of Design, 1979.
- [66] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.