Enhancing Autonomous Access Hole Detection

Ben Waismark, Alexander Ferworn, and Jimmy Tran
Ryerson University
Toronto, Ontario, Canada

***Abstract* – When buildings collapse, the resulting rubble may trap live victims underneath it, entombing them. Voids beneath the rubble may allow victims to stay alive. However, chances of survival dramatically decrease following the initial 24 hours after a disaster occurs. Hence, first responders must act quickly to rescue live victims. The work presented showcases enhancements to an autonomous access hole detection system, considerably increasing its accuracy.**

Keywords - Urban Search And Rescue; USAR; Disaster; Rubble; Access Hole

I. Introduction

Large emergencies that are beyond the ability of local emergency services to cope with, requiring significant extra-jurisdictional resources, are referred to as disasters. Disasters typically pose threat to large numbers of people in an area where they occur. Rubble of collapsed buildings may crush or entomb live victims inside voids that form in the rubble. Voids are more likely to form in rubble of collapsed 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 disasters strike urban areas, entrapping victims under rubble, Urban Search And Rescue (USAR) teams are typically responsible for the ensuing response operations.

When emergency situations result in collapsed structures entombing victims in voids, a search for access holes into the rubble and for signs of life must be completed before rescue operations proceed. This search process is performed by trained search and rescue teams; 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 are used to locate trapped and hidden live victims.

Locating trapped victims and access holes 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.

The work presented improves and extends previous work by Kong [2] – where access holes are functionally defined using a set of visual cues and geometric constraints. Colour-depth imagery is used as input for the access hole detection algorithm. The input frames are segmented into superpixels based on the depth image. The resulting segmentation is used in the scoring process for both the depth and colour images. Each of the superpixels are scored on selected attributes and lumped into a final access hole likelihood score.

The developed extension allows the system to process a continuous stream of frames; the stream of frames aid the accurate recognition of access holes. Multiple frames allow the system to make decisions based on multiple views of a region to provide better recognition. A revised segmentation and depth interpolation method is introduced, providing a more robust process, which helps mitigate corrupt and incomplete depth readings. These additions improve the accuracy of the access hole detection algorithm, providing a reliable system for location and localization of access holes.

II. Technical Approach

A functional definition given by Kong [2] is used: An access hole in an USAR environment is defined as such a hole that allows access to an adult human searcher, with three attributes to define an access hole: "(i) depth disparity, (ii) hole size and (iii) photometric brightness."

The input dataset is processed through several steps to achieve a final detection. The steps include segmentation, depth interpolation, attribute scoring,

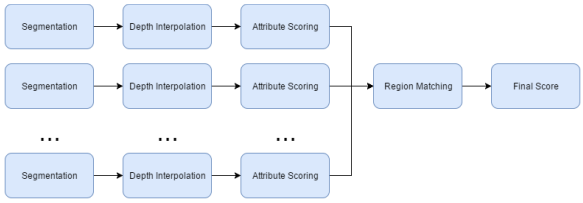

Fig. 1. The detection pipeline.

matching segments across frames and computing mean score for every group of segments. A diagram describing this flow is provided in Figure 1.

The first step in the detection pipeline is superpixel segmentation – where input colour and depth frames are segmented. Ideally, every segment represents a single entity, which can be an object or a hole. The colour and depth images are each initially segmented into N segments. Neighbouring segments in the colour image segmentation are joined together if their brightness is below a threshold. The average luminosity of a segment, extracted using the YUV colour space, is used to gauge brightness throughout this process. A plane of best fit is fitted to each of the superpixels generated by segmenting the depth frame. Neighbouring segments in the depth segmentation are joined together if their planes of best fit are similar. 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; if the points fall closely together, below the error threshold, the segments are joined. The resulting segmentations of the colour and depth images are overlaid, producing a segmentation that is based on both the colour and depth components; in this segmentation, neighbouring segments are joined together by comparing the similarity of the segments' planes of best fit. This segmentation process introduces a reliance on both the colour and depth components, while the previous work relied on a segmentation achieved by segmenting the depth frame only.

The second step is the interpolation of depth data within each segment. Using a plane of best fit for each segment, depth data is interpolated. This process mitigates missing and corrupt depth data. The approach used is similar to the depth interpolation process described in [3]. Based on the assumption that objects and surfaces are generally composed of flat regions, depth points are projected from a flat plane of best fit. Linear interpolation is preferred for two main reasons: (i) Performance - fitting linear planes is computationally less expensive than fitting curved surfaces 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 receive a higher score. The scores of all attributes are combined to form the final access hole score for every segment. The attributes and scoring process are identical to those used in the original algorithm by Kong [2].

Lastly, for each of the potential holes detected, adjacent frames are searched for visually similar segments. For each segment, features are extracted and compared against features of segments in other frames; the current implementation uses minimum Eigenvalue features [4]. Scores for groups of similar looking segments across multiple frames, assumed to contain the same object (hole), are averaged to provide a more accurate score. The averaged score over a sequence of frames allows multiple points of view to be used. This helps reduce the total number of output detections by removing duplicate outputs. Because scores of false detections are averaged, false positives and false negatives are also reduced by using this process. Holes with a high score are considered detected holes.

III. Evaluation

The work presented was tested and its performance was compared to the results of the initial

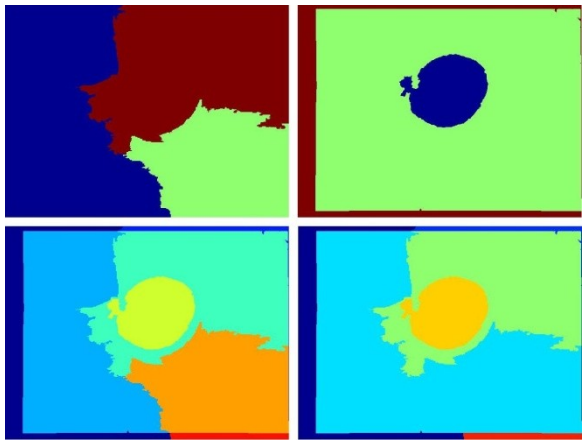

Fig. 2. Segmentation of colour image (top left), depth image (top right), overlaid segmentations (bottom left) and overlaid segmentations with merged segments (bottom right).

Fig. 3. Sample output detections marked by bounding boxes.

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 1: AP for several numbers of segments used during initial segmentation of the colour and depth images.

access hole detection system introduced by Kong [2]. To allow for an accurate comparison with results of the previous work, parameter values were set to the same values as in Kong's work. The evaluations presented make use of the dataset introduced by Kong [2]. The 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 an unmanned aerial vehicle (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.

Multiple evaluations showcasing the performance and accuracy of the proposed work are presented. First, a direct comparison with Kong's work is presented, based solely on individual segment scores, without the use of the multi-frame extension. The algorithm's performance is measured using Precision-Recall (P-R) [5]. 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 at least 50% of a ground truth bounding box. Table 1 shows a measure of Average Precision (AP) based on the number of segments used during segmentation. The current implementation uses Entropy Rate Superpixel (ERS) [6] segmentation to segment both colour and depth frames. The same superpixel segmentation algorithm was used throughout Kong's work.

Although based on the information in Table 1, it may seem that a lower number of segments should be preferable, segmentations using less than 7 segments failed to achieve a 100% recall. Based on this information, all images were segmented into 8 segments for the evaluation of single frame performance. Table 2 shows the APs of photometric, geometric and combined attributes.

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 approach, 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, the P-R curve with and without multi-frame matching is displayed. Table 3 displays the precision achieved for a 100% recall rate as well as the AP with and without the use of multi-frame scoring, for segmentations into 5 and 8 superpixels. 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 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

	Geometric	Photometric	Combined
Average Precision	0.37	0.45	0.47

Table 2: Geometric and Photometric attributes' AP

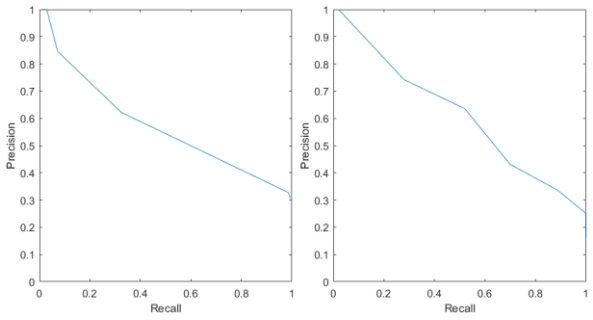

Fig. 4. Precision-Recall curves with multi-frame matching (left) and without multi-frame matching (right).

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

levels.

The main goal of the work presented is to provide a more accurate tool for first responders. A notable improvement has been achieved; the highest AP reported by Kong is 0.43, the work presented achieves an AP of 0.65 as demonstrated in Table 1 – an improvement of 22%.

The highest AP was obtained with the number of segments for initial segmentation set to 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.

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 reach an even higher combined AP of 0.47.

Number of Segments	5		8	
Multi-Frame Scoring	Yes	No	Yes	No
Output Detections	758	2464	1102	4107

Table 4: Number of possible holes for verification

The AP curve resulting from the use of multi-frame scoring is shown in Figure 4. 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 3 and Table 4. 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.

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] C. Kong, "Discovering access holes in disaster rubble with functional and photometric attributes," Ryerson University, 2015.

[3] 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.

[4] 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.

[5] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.

[6] 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.