2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Finding Disaster Victims: A Sensory System for Robot-Assisted 3D Mapping of Urban Search and Rescue Environments

Zhe Zhang, Hong Guo, Goldie Nejat, and Peisen Huang
Department of Mechanical Engineering, State University of New York (SUNY) at Stony Brook
Stony Brook, NY, USA, 11794-2300, Email: zhzzhang@ic.sunysb.edu, Goldie.Nejat@stonybrook.edu

Abstract – **In this paper the first application of utilizing a unique 3D *real-time* mapping sensor for sequential 3D map building within a Visual Simultaneous Localization and Mapping (SLAM) framework in unknown cluttered urban search and rescue (USAR) environments is proposed. The sensor utilizes a digital fringe projection and phase shifting technique to provide *real-time* 2D and 3D sensory information of the environment. The proposed sensor is unique over current technologies, in that it can directly map rubble in 3D and in *real-time* at a frame rate of up to 60 fps. Furthermore, we propose the development of a novel 3D Visual SLAM method utilizing both 2D and 3D images taken by the sensor for robust and reliable landmark identification, mapping and localization algorithms utilizing a Scale Invariant Feature Transform (SIFT) -based approach. Preliminary experiments show the potential of the proposed 3D real-time sensory system for such unknown cluttered USAR environments.**

Index Terms – Urban search and rescue, structured light, 3D Visual SLAM, SIFT.

I. Introduction

The catastrophic earthquakes that hit northern and southern California in 1989 and 1994, and Kobe, Japan in 1995 and the terrorist attacks on the World Trade Centers in 2001 have clearly demonstrated the need for specially trained resources to respond to incidents of partial or complete structural collapse caused by these types of major disasters. Urban search and rescue (USAR) is defined to be the emergency response function which deals with the collapse of man-made structures [1].

With the advancement of robotic research in recent years, rescue robots have been developed to address this particular conundrum and to lessen the burden on rescue workers. However, there are a number of challenges that roboticists must face in designing a USAR robot. In particular, major advances are needed in sensor techniques and sensor information interpretation for two main tasks: (i) victim identification, and (ii) navigation of the robot. Most robots' relationship to their environments is limited by sensor technologies and cost, where their location in the environment, the layout of the environment, and the presence of victims is usually extracted from a single 2D video camera [1]. Furthermore, all robots that operate in USAR environments do not have any a priori information about landmarks in the scene and due to the nature of the surroundings cannot employ GPS. A robot operator in USAR environments faces the important tasks of remembering, recognizing and diagnosing a scene and how the robot and its camera are positioned and oriented within the scene merely from this camera. Often times, this leads to disorientation, the robot getting stuck, and not being able to identify victims that are present in the scene. In order to address the limitations of current sensors utilized in USAR, we propose the development of a 3D mapping sensory system for the effective 3D mapping of USAR environments and localization of mobile robots to minimize the stress and burden on the operator.

This paper presents a major effort in developing a compact 3D sensory system for robotic search and rescue operations in unknown chaotic environments. We describe the first application of using a structured light sensor for sequential map building within a 3D Visual Simultaneous Localization and Mapping (SLAM) framework. The proposed sensory system is a unique cost-effective solution. Its main advantages over current technologies, is that it can directly map rubble in 3D and in *real-time* at a frame rate of up to 60 fps. With this sensor, 3D images along with 2D images of the rubble surrounding the robot can be made available to the operator in *real-time*. The performance of the 3D real-time mapping sensor will be independent of three main limiting factors of current sensors: (i) the use of a scanning mechanism, which is time-consuming in real-time applications, (ii) slow scanning speed; our sensor can provide 3D mapping in real-time, and (iii) the illumination conditions of the environment; our sensor will successfully work in dim lit and dark environments. The proposed system will map the 3D environment fast and reliably without the need for human intervention. The three main conundrums that will be addressed in order to generate an accurate 3D map of the environment are: (i) acquirement of 3D information about the landmarks in the scene, (ii) landmark identification, and matching, and (iii) 3D SLAM.

In Section II, we review the current sensors utilized for mapping and define the SLAM problem. In Section III, we outline the overall 3D real-time mapping sensory system architecture. Experimental results are shown in Section IV.

II. Sensory Systems for Mapping USAR Environments

A. Current Sensors for Mapping

Among all the existing sensory techniques that can be potentially used for mapping, stereovision is probably the most studied method. A stereo camera is the prime example of a passive optical triangulation system. Traditional stereovision methods estimate shape by establishing spatial correspondence of pixels in a pair of stereo images. Determining the correspondences between left and right

view by means of image matching, however, is a slow process. Furthermore, for 3D reconstruction, passive stereovision techniques depend heavily on cooperative surfaces, mainly on the presence of surface textures, and on ambient light [2]. Such texturing is absent in USAR environments where the surroundings are dark and covered in gray dust. Recently, Zhang et al., [3] developed a new concept called spacetime stereo, which extends the matching of stereo images into the time domain, where the shortcoming again is the requirement of the time-consuming task of matching of stereo images. Therefore, it is difficult to reconstruct high-resolution 3D shapes from stereo images in *real-time*. 3D cameras have also been proposed for mapping. However, the pixel array size of these cameras is limited and hence the resolution of both the 2D and 3D images can be low.

Laser scanning consists of using a laser light source that sweeps a thin laser stripe across a scene. Simultaneously, a light sensor, i.e. camera, acquires the scene, where the surface of the scene is measured via triangulation, or time-of-flight. The main disadvantage of laser scanners for robotic 3D mapping of USAR environments is that they require a lot of time for scanning, due to the fact that the laser stripe has to be physically moved across the scene to digitize the surface, and hence cannot provide real-time range data acquisition. Other disadvantages of laser scanners are high cost of production of their hardware components (i.e., costs are in the range of several tens of thousands), they are bulky and heavy for a small robot, and they can produce a variety of wrong points in the vicinity of edges.

Due to the aforementioned limitations of laser scanning and stereovision for mapping in cluttered USAR environments, a hand full of research projects have been proposed for the development of unique sensors for such environments. In [4], Kurisu et al. proposed the use of two different laser range finders for 3D mapping of rubble: (i) a ring of laser beam module and an omnivision CCD camera, (ii) and an infrared laser module with a CCD camera to capture the laser image and another camera for capturing the texture. The optimal range of this system was determined to be 300 mm. There are two main limitations to these types of sensors: (i) they do not address real-time range data acquisition, and (ii) their reliance on robot internal sensors for mapping, in particular they can only measure in the x,y plane, the z-direction measurement for the 3D information is based on the robot's inaccurate internal sensors.

B. The SLAM Problem

In order to map its environment, the robot must be able to determine where it is in relation to its surroundings. Due to the increase in uncertainty over time, robot sensors such as odometers are not sufficient for such a task. In indoor environments, usually the robot is mapping scenes in which known landmarks exist; hence the location of these landmarks can be utilized in order to localize the robot. In outdoor environments, accurate sensors such as GPS can be utilized to determine the location of the robot. However, all robots that operate in USAR environments do not have any a priori information about landmarks in the scene and cannot employ GPS/radio positioning due to the nature of the surroundings (i.e., inside cluttered collapsed buildings). Furthermore, USAR environments even more unique due to the characteristics of the uneven terrain. Hence, a localization algorithm is crucial while mapping the unknown site.

Some attempts have been made to directly formulate SLAM for rescue environments [5,6]. In [5], Ishida et al. utilized a 2D laser scan matching-based SLAM method to explore an environment with flat ceilings. A global map is then created using several sphere digital elevation maps (SDEM) and the relative locations among them. But the robot's yaw orientation is extremely difficult to estimate by this method and it leads to errors in localization. Furthermore, the assumption of the environment having a flat ceiling limits the method's application. In [6], Yokokohji et al. have conducted some preliminary work on 3D SLAM assuming known data association. Based on robot accelerations and 2D range measurements from a laser range finder, an Extended Kalman Filter (EKF) is utilized for system state estimation. Only 2D positions of the landmarks can be measured, thus, relying on inaccurate robot sensory information for the third coordinate.

Only recently interest has increased in utilizing cameras for SLAM applications, known as Visual SLAM [7]. Cameras are affordable and compact, and can be used to provide 3D range information. Furthermore, they have a high rate of acquisition and high angular resolution.

Recently, attempts have also been made in the literature to develop methods for identifying distinctive invariant features from images that can be used to perform matching of objects from different views. One particular method is the Scale Invariant Feature Transform (SIFT) developed by Lowe in [8]. This approach transforms an image into a large collection of local feature vectors, each of which is invariant to image translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. The resulting feature vectors are called SIFT keys. This method has been utilized effectively on 2D grayscale images to identify and match invariant features. Moreover, it works efficiently for object recognition problems where a training image of the object of interest is given.

For visual SLAM in USAR environments we propose the utilization of SIFT features for identifying and matching of non a priori landmarks. However, since visual systems are restricted to the sensing and processing of information which can be displayed as 2D projections, we propose the utilization of 3D depth images of the scene. This additional information is utilized to extract strong evidence in discontinuity between multiple objects detected in a scene in order to locate large distinguishable landmarks in a USAR environment for 3D mapping of the environment.

III. 3D MAPPING SYSTEM ARCHITECTURE

In this section the main components of the proposed 3D real-time sensory system are described.

A. Real-Time 3D Mapping Sensor

A major group of 3D scanning techniques is structured light, which includes various coding methods and employs varying number of coded patterns [9]. Structured light systems are commonly adopted because they are simple for recognition, sampling, modelling, and coordinate calculation. Unlike stereovision methods, structured light methods usually use simpler processing algorithms. Unlike its laser counterparts, capturing happens in a single step, allowing the entire surface to be digitized by a single acquisition. Therefore, it is more likely to achieve real-time performance. Many techniques in this group have been developed, in particular using a single color pattern to boost the speed [10-13]. Even though it shows potentials for real-time 3D scanning, scanning results are affected to varying degrees by the variations of the object surface color. Others use multiple coded patterns but switch them rapidly so that they can be captured in a short period of time. However, the need to switch the patterns by repeatedly loading patterns to the projector limits scanning speed.

Huang et al. recently proposed a high-speed 3D shape measurement technique based on a digital fringe projection and phase shifting technique, which uses three phase-shifted, sinusoidal grayscale fringe patterns to provide pixel-level resolution [14,15]. The patterns are projected to the object with a switching speed of 360 fps. This technique is precisely what is needed to overcome the limitations of the aforementioned mapping systems currently in use, but little work has been done to apply this technology to the mapping and localization problem. Herein, we propose the development of a real-time 3D mapping sensor based on the basic concept of the digital fringe projection and phase shifting technique for mapping of USAR environments, Fig. 1. Firstly, the technique consists of generating a color fringe pattern with its red, green, and blue channels coded with three different patterns created by a PC. When this pattern is sent to a Digital-Light-Processing (DLP) projector working in black-and-white (B/W) mode, the projector projects the three colour channels (R,G,B) in sequence repeatedly and rapidly over the scene. As a result, three greyscale fringe patterns with phase shift are projected onto the scene sequentially. A B/W high speed CCD camera synchronized with the projector captures the scene image consecutively.

The fundamental concept behind this structured light 3D measurement system is PSI (Phase Shift Interferometry). The light intensity of an arbitrary point (x, y) in images captured with patterns (I_1, I_2, I_3) can be expressed with Eq. (1)-(3) respectively:

$$I_1(x,y) = I'(x,y) + I''(x,y) * \cos[\phi(x,y) - \alpha], \quad (1)$$

$$I_2(x,y) = I'(x,y) + I''(x,y) * \cos[\phi(x,y)], \quad (2)$$

$$I_3(x,y) = I'(x,y) + I''(x,y) * \cos[\phi(x,y) + \alpha], \quad (3)$$

where $I'(x,y)$ is the average intensity; $I''(x,y)$ is the intensity modulation; $\phi(x,y)$ is the unknown phase at point (x,y); and α is the constant $2\pi/3$. Three unknowns $I'(x,y)$ $I''(x,y)$ and $\phi(x,y)$ can be solved with the above equations. Once the phase information $\phi(x,y)$ is retrieved, the 3D information of the scene could be reconstructed by applying both a phase unwrapping algorithm and a triangulation algorithm. Meanwhile, the texture information of the landmarks could be easily retrieved from every 3 consecutive 2D fringe images, and mapped onto a landmark's 3D model. Due to the fringe pattern projected by the projector having a high brightness, the system is more robust to environmental noises than those using stereovision methods.

Fig. 1 shows the real-time structured light 3D shape measurement system setup. A DLP projector projects fringe patterns with the frequency of 360Hz. The B/W high speed CCD camera synchronized with the DLP captures the fringe images at the frequency of 90Hz. Based on the above PSI technique, each frame of the 3D shape is reconstructed using three consecutive fringe images. Therefore, the 3D data acquisition speed of the system is 30 frames per second. Together with the fast 3D reconstruction algorithms and parallel processing software we have developed, high-resolution real-time 3D shape measurement is realized at a frame rate of up to 30 3D frames per second and a resolution of 532×500 points per frame.

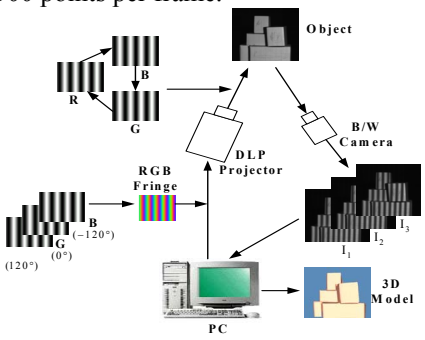

Fig. 1: System Diagram.

B. Identifying Landmarks for 3D Visual SLAM

When traveling in 3D cluttered environments, data association (i.e., landmark identification and matching) becomes a pertinent problem. In USAR environments there could exist many repetitive features. As the robot moves, it must be able to determine whether different sensor measurements correspond to the exact same landmark in its environment. In most cases, the SLAM problem has been addressed under known data association [16]. However, in most situations including USAR environments, this is not the case. Furthermore, incorrect data association can induce extreme errors in SLAM solutions. By utilizing a SIFT-based approach and incorporating 3D grayscale depth imagery, we will be able to use more reliable and robust recognition and matching between landmarks from different images, therefore minimizing false matches. If an object in the foreground of an image is similar in intensity to the background, it is difficult to determine its boundaries. The use of depth images solves this problem, since a foreground object will always be at a closer depth, and can therefore be easily detected and identified as a potential landmark. The

SIFT approach consists of four main stages [17]: (i) Scale-Space Extrema Detection, (ii) Keypoint Localization, (iii) Orientation Assignment and (iv) Keypoint Descriptor Assignment. Preliminary work, conducted by Nejat et al, has included the development of a Nearest Neighbour (NN) 3D keypoint search method and Canny-Deriche edge detection for landmark identification [18, 19].

The overall proposed method will be discussed herein outlining its most pertinent stages: (i) identifying keypoints, (ii) identifying clusters, and (iii) matching of clusters.

1) Keypoint Identification: The keypoints of an image and their dimensional descriptors are determined (and stored) by finding the keypoints and descriptors for the (i) 2D image, and (ii) corresponding 3D depth image utilizing the four stages of the SIFT algorithm, i.e., Table 1. In general due to shadowing effects and texture changes, a number of keypoints can be identified in the 2D images. Fig. 2(a) shows keypoints (green circles) found on a box, with multiple keypoints on the flat surfaces of the box. In the 3D (i.e., depth) image, Fig. 2(b), the keypoints on the flat surfaces are no longer present since there is no significant change in depth on these surfaces. We can analyze and cluster the keypoints we found in the 2D image based on the keypoints found in the 3D image in which for the latter image shadowing and texture effects are not present.

2) Keypoint Clustering: Clustering keypoints is not only important in defining landmarks but also in reducing the number of keypoints of interest. The 2D and 3D images have a one-to-one correspondence. Mainly, if a keypoint does not exist in the same pixel in the 3D image, then the keypoint is assumed to be due to image shadowing and texture effects. Clusters are bound in regions where a large number of keypoints in the 3D image do not exist, i.e., they have considerably the same depth information. These clusters can be used to represent large distinguishable landmarks in the scene. Hence, we can identify a cluster of keypoints in the 2D image by bounding them by keypoints in the 3D image. We will use the clustering technique we developed in [19] to do this.

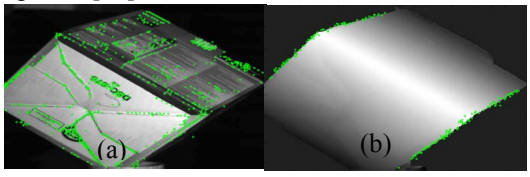

Fig. 2: (a) 2D image, (b) 3D image of landmarks.

3D Image Analysis

Keypoints that are determined in the 3D image are grouped together based on grayscale depth information into depth clusters, where they represent the cluster boundaries for the keypoints in the 2D image. The depth grayscale is determined from 0 to 255.

Step 1: Initially, each keypoint is specified by 5 parameters: x location, y location, depth, scale and orientation, and stored in the matrix $\mathbf{A}_{ln}$, where l represents the number of keypoints and n represents the number of parameters, i.e., Table 1.

Table 1: Step 1 of algorithm: Keypoint parameters matrix **A**.

Keypoint	x position	y position	Depth	Scale	Orientation
1	80.13	259.74	162	27.14	-1.357
2	373.37	115.63	123	18.89	-1.751

Step 2: The Canny-Deriche edge detection algorithm is used to determine potential boundaries of objects in the scene by identifying edge pixels via gradient intensity [20]. Fig. 3(a) and (b) show the 3D image of a scene and object boundaries obtained using the Canny-Deriche method. In relation to other edge detection algorithms, the Canny-Deriche method has shown to be the most optimal for our work.

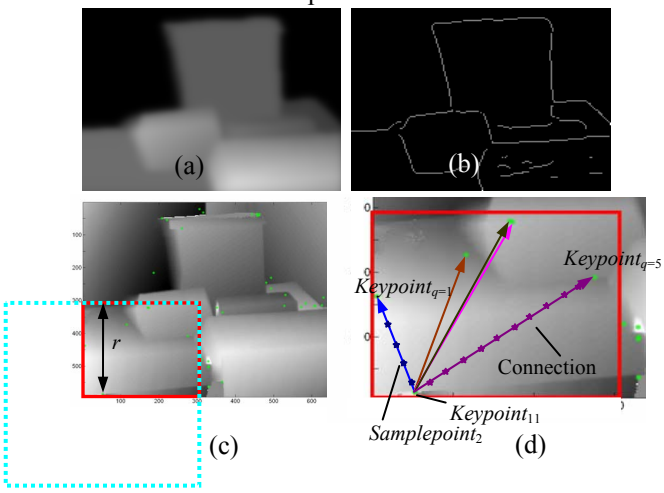

Fig. 3: Step 2: (a) 3D image of scene, (b) Canny-Deriche algorithm to define boundaries for objects, (c) Step 3 of algorithm: Search region, (d) Step 4 of algorithm: Finding the NN keypoint.

Step 3: The NN search method starts from an arbitrary corner of the image and locates the closest keypoint ($keypoint_{jk}$, where j=1 represents cluster number and k=1 represents keypoint number in the cluster) to the corner. A square of side length $2r$ is drawn symmetrically around the keypoint to search for its NN keypoints, Fig. 3(c). If no keypoints are initially found, r is continuously incremented until keypoints are detected. Each detected keypoint and its 5 parameters are stored in a temporary matrix $\mathbf{B}_{pn}$ for evaluation, where p represents the number of detected keypoints. For the initial point, only the portion of the square that encompasses the image is searched, i.e., the red square in Fig. 3(d).

Step 4: A vector is drawn from the initial keypoint, $keypoint_{11}$ to every keypoint in matrix **B**. N number of points on each vector are sampled for depth information, $samplepoint_i$, where i=1,2,…,N, Fig. 3(d). The NN keypoint, $keypoint_{12}$ is determined to be the keypoint with the minimum change in depth information from $keypoint_{11}$ and whose corresponding sample points have the smallest variation in depth from itself (i.e. $keypoint_q$, where q=1,…,p-1,p) and $keypoint_{11}$:

$$Minimum\_depth\_value(Maximum\_depth\_value) = \min(\max)[\mathbf{A}(keypoint_{11},3), \mathbf{B}(keypoint_q,3)] \quad (4)$$

$$Minimum\_depth\_value - threshold < samplepoint_i\_depth < Maximum\_depth\_value + threshold \quad (5)$$

The objective of sampling multiple points between the keypoints is to ensure that boundaries of objects are not crossed.

Steps 3 and 4 are repeated until all keypoints in the corresponding cluster are identified. The sample points from previous keypoints in the cluster are stored with their corresponding keypoints. This information is used along with sample points determined for the keypoint of interest in deciding whether the keypoint belongs to the cluster and its order within the cluster:

$$keypoint_{j(k+1)} = f[(samplepoint_i)_m, (samplepoint_i)_{k+1}, keypoint_{jm}], \quad (6)$$

where $m = 1,\ldots, k\text{-}1, k$. For every keypoint that is added to the cluster, its **A** matrix information is updated with the following additional parameters: order in the cluster, number of connections to other keypoints, depth information stored from sample points. In order for a keypoint to be considered in the cluster, it must have a minimum of two connections to other keypoints in the cluster.

Once all keypoints are determined in a particular cluster, a new matrix with all the corresponding keypoint information is defined for that cluster. The algorithm, then, searches for new clusters starting at other corners of the image and Steps 2-4 are repeated accordingly. Once completed, the algorithm defines the keypoints that have not been clustered and determines if they belong to an old cluster or will create a new cluster.

2D Image Analysis

Once all depth clusters in the 3D image have been identified, they can be used to identify their corresponding keypoints in the 2D image. Each depth cluster represents the boundary conditions for the 2D keypoints. Since there exists a one-to-one correspondence between the 3D and 2D images, the boundaries can be superimposed on the 2D image. Herein, cluster boundaries are represented by the connection vectors between the keypoints in the depth clusters. Based on the pixel occupancy of the boundaries, 2D keypoints that are located within these boundaries are identified and stored in the cluster matrix. Each cluster is defined to represent a landmark in the environment, Fig. 4. It is important to note that this clustering method does not attempt to represent the shape of the landmark in the environment; it merely identifies detectable regions that can represent a portion of a true landmark and that can be matched in successive images with different viewpoints.

Matching of Clusters

Matching of clusters relies on finding the same clusters in consecutive images by matching keypoints from the clusters from previous frames with ones in the new cluster of the current frame. We utilize the Best-Bin-First (BBF) method proposed by Lowe, in [17]. Herein, key descriptors of the keypoints are matched which can correspond to finding a set of NNs to a query point. The advantage of this method is its ability to handle high-dimensional spaces, i.e., the 128 dimensional descriptor vectors. Fig. 4(c) illustrates matching between two clusters determined in two different viewpoints of a scene. The blue lines represent the keypoints that were matched within the two images. The effectiveness of the method is shown in Fig. 4(c), where the majority of the matches made are correct matches.

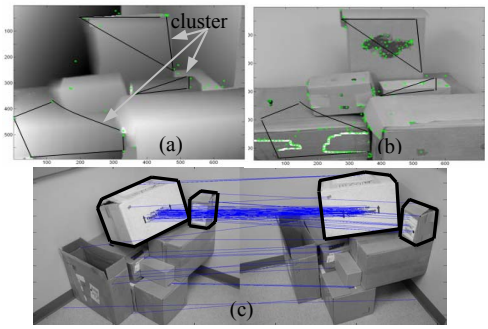

Fig. 4: Cluster results: (a) 3D image, and (b) 2D image, and (c) Matching of clusters in different images.

C. Visual SLAM

The 3D range information of the landmarks is provided by the 3D sensor via a point cloud with respect to the camera coordinate frame. This information corresponds to the pixels the landmark occupies in the 2D image. Hence, by identifying the location of the SIFT keypoints representing one landmark in the 2D image, its 3D range information in the camera coordinate frame can be determined. The 3D coordinates of the same SIFT keypoints (SIFT pairs) in different images can be utilized to determine the 6 DOF ego-motion parameters (i.e., ΔX, ΔY, ΔZ, $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$). At least three pairs of SIFT keypoints are needed to estimate the ego-motion transformation. Since the position of the camera relative to the robot's coordinate frame is known, the transformation between the robot at two locations can be determined. By utilizing this information and the localization information from the previous position, the robot's location can be estimated. Furthermore, once the alignment of the same landmarks is determined between different visual sensor locations, the corresponding 3D range information of the scene can be stitched together for reconstruction of the USAR environment via a 3D global map.

IV. Experiments

Several preliminary experiments were conducted to verify the proposed sensory system consisting of the PLUS U5-632 Digital projector with 1024×768 resolution and 3000 lumens light output and the Dalsa CA-D6-0512 B/W high speed CCD camera (resolution 532×500). The system was placed on top of an all-terrain robot, Fig. 5. While the robot navigated through an environment filled with brown cardboard boxes, 2D and 3D images were taken in real-time. The effective range of measurement of the system was 0.7~1.4m, with the current lens configuration of the camera and projector. The brown cardboard boxes mimic a USAR environment in the sense that they represent different shapes of objects and also the small variation in color of the scene.

Utilizing the images taken by the sensor, the landmark identification and matching, and Visual SLAM algorithms were implemented.

On average 264 and 1073 keypoints were determined in the 3D and 2D images, respectively. The maximum allowable change in depth between keypoints in a cluster was defined to be 80 (on a scale from 0 to 255), and the depth threshold for the sample points in Eq. (5) was set to 20. In the experiment shown in Fig. 5, 8 clusters were found and matched at two different robot poses. The matched keypoint pairs and their corresponding 3D range information were utilized to estimate the 6 DOF ego-motion via the Levenberg-Marquadt nonlinear solver: i.e., ΔX=-69.85cm, ΔY=-9.99cm, ΔZ=4.99cm, $\Delta\alpha$=-4.76°, $\Delta\beta$=3.02°, $\Delta\gamma$=-4.71°. The true ego-motion determined by a high-precision motion control system was: ΔX=-70cm, ΔY=-10cm, ΔZ=5cm, $\Delta\alpha$=-5°, $\Delta\beta$=3.18°, $\Delta\gamma$=-4.87°. This ego-motion estimation can be utilized for: (i) alignment in the stitching of 3D range information from the images, and (ii) to globally localize the robot in USAR environments. The proposed method took approximately 10s on a Pentium IV 3.0GHz 1Gb of RAM system.

V. Conclusions

In this paper, we propose the first application of using a structured light sensor for sequential map building within a 3D SLAM framework. A SIFT-based method is utilized to analysis the 2D and 3D images taken by the sensor for effectively identifying large distinguishable landmarks in the scene and matching the landmarks for 3D Visual SLAM in a USAR environment. The preliminary experiments show the potential of the proposed method for such applications.

References

[1] R. R. Murphy, "Human-Robot Interaction in Rescue Robotics," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 34, No. 2, pp. 138-153, 2004.

[2] S. Se and P. Jasiobedzki, "Photo-realistic 3D Model Reconstruction", *IEEE Int. Conference on Robotics and Automation (ICRA)*, pp. 3076-3082, 2006.

[3] L. Zhang, B. Curless, and S. Seitz, "Spacetime Stereo: Shape Recovery for Dynamic Senses," *Computer Vision and Pattern Recognition*, 2003.

[4] M. Kurisu, Y. Yokokohji, and Y. Oosato, "Development of a Laser Range Finder for 3D Map-Building in Rubble The 2nd Report: Development of the 2nd Prototype," *IEEE Int. Conference on Mechatronics and Automation*, pp. 1842-1847, 2005.

[5] H. Ishida, K. Nagatani, Y. Tanaska, "Three-Dimensional Localization and Mapping for a Crawler-type Mobile Robot in an Occluded Area Using the Scan Matching Method," *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, pp. 449-454, 2004.

[6] Y. Yokokohji, M. Kurisu, S. Takao, Y. Kudo, K. Hayashi, and T. Yoshikawa, "Constructing a 3-D Map of Rubble by Teleoperated Mobile Robots with a Motion Canceling Camera System," *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, pp. 3118-3125, 2003.

[7] A. J. Davison, and D. W. Murray, "Simultaneous Localization and Map- Building Using Active Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v 24, n 7, pp. 865-880, 2002.

[8] D. G. Lowe, "Object Recognition from Local Scale-invariant Features," *Int. Conference on Computer Vision,* pp. 1150-1157, 1999.

[9] K. G. Harding, "Color Encoded Moiré contouring," *SPIE*, Vol. 1005, pp. 169-178, 1988.

[10] J. Salvi, J. Pages, and J. Batlle, "Pattern Codification Strategies in Structured Light Systems," *Pattern Recognition,* Vol. 37, No. 4, pp. 827-849, 2004.

[11] Z. J. Geng, "Rainbow 3-D Camera: New Concept of High-speed Three Vision System," *Opt. Eng.* 35, pp. 376-383, 1996.

[12] P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang, "Color-Encoded Digital Fringe Projection Technique for High-Speed Three-Dimensional Surface Contouring," *Opt. Eng.* 38, pp. 1065-1071, 1999.

[13] L. Zhang, B. Curless, and S. M. Seitz, "Rapid Shape Acquisition using Color Structured Light and Multi-pass Dynamic Programming," *IEEE Int. Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 24-36, 2002.

[14] S. Zhang and P. Huang, "High-Resolution, Real-time 3D Shape Acquisition," *IEEE Computer Vision and Pattern Recognition Workshop on Realtime 3D Sensors and Their Uses*, Vol. 3, pp. 28–37, 2004.

[15] S. Zhang and P. S. Huang, "High-resolution, real-time 3-D shape measurement," *Opt. Eng.*, Vol. 45, 2006, In print.

[16] R. Smith, M. Self, and P. Cheeseman, "Estimating Uncertain Spatial Relationships in Robotics," Autonomous Robot Vehicles, pp. 167–193. Springer, 1990.

[17] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[18] G. Nejat and Z. Zhang, "The Hunt for Survivors: Identifying Landmarks for 3D Mapping of Urban Search and Rescue Environments," *The World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2006)*, Vol. 2, pp. 252-256, 2006.

[19] G. Nejat and Z. Zhang, "Finding Disaster Victims: Robot-Assisted 3D Mapping of Urban Search and Rescue Environments via Landmark Identification", *IEEE Int. Conference on Control, Automation, Robotics and Vision (ICARCV 2006)*, Singapore, pp. 1381-1386, 2006.

[20] Biomedical Imaging Group, "Edge Detector", 2003, Available HTTP: http://bigwww.epfl.ch/demo/jedgedetector/.

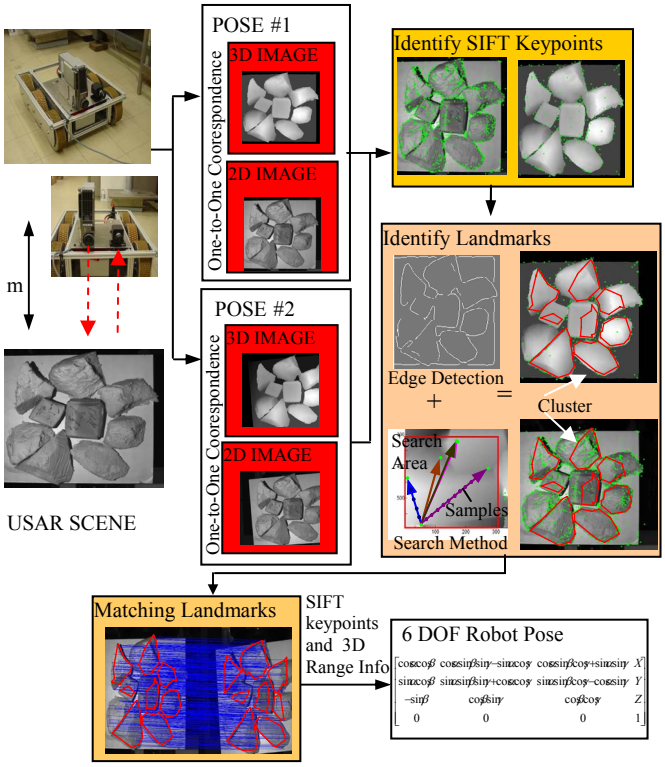

Fig. 5: Experimental results.