This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

A novel 3D sensory system for robot-assisted mapping of cluttered urban search and rescue environments

Zhe Zhang · Goldie Nejat · Hong Guo · Peisen Huang

Abstract In this paper, the first application of utilizing a unique 3D sensor for sequential 3D map building 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 high-resolution 3D information of rubble filled environments can be acquired from the single sensor at a speed of 30 frames per second (fps). Furthermore, we propose the development of a novel robust and reliable landmark identification technique that utilizes both 2D and 3D depth images taken by the sensor for 3D mapping. Preliminary experiments show the potential of the real-time 3D sensory system and landmark identification scheme for robotic 3D mapping in unknown cluttered USAR-like environments.

Z. Zhang
Autonomous Systems Laboratory, Department of Mechanical Engineering, The State University of New York at Stony Brook, Stony Brook, NY 11794-2300, USA
e-mail: Zhe.Zhang@sunysb.edu

G. Nejat (✉)
Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto M5S 3G8, Canada
e-mail: nejat@mie.utoronto.ca

G. Nejat · H. Guo · P. Huang
Department of Mechanical Engineering, The State University of New York at Stony Brook, Stony Brook, NY 11794-2300, USA

H. Guo
e-mail: hguo@ic.sunysb.edu

P. Huang
e-mail: peisen.huang@stonybrook.edu

Keywords Urban search and rescue · Structured light sensing · Landmark identification · 3D mapping

1 Introduction

The catastrophic earthquakes that hit Haiti and Chile in 2010, northern and southern California in 1989 and 1994, Kobe Japan in 1995 and the Izmit region in Turkey in 1999, 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 [20]. In both human caused and natural disasters, the fundamental tasks at hand are (i) to find and rescue victims in the rubble or debris as efficiently and safely as possible, and (ii) not to further endanger the survivors or put human rescue workers' lives at great risk.

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 the rescue workers. 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 [20]. 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. To address the limitations of current sensors utilized in USAR, we propose the development of a 3D sensory system

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

for the effective 3D mapping of USAR environments to minimize the stress and burden on the operator. We describe the *first application* and *feasibility study* of a structured light sensor for sequential map building in such environments. Our proposed sensory system is a unique cost-effective ($5,000) solution. It can directly provide both 3D and 2D sensory information in real time at a frame rate of 30 fps with pixel-level resolution. The real-time capabilities of the proposed sensory system can be a great asset in time sensitive USAR environments. The performance of the 3D real-time 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 sensory information in real-time, and (iii) the illumination conditions of the environment; our sensor will successfully work in dim lit and dark environments. The two main conundrums that will be addressed to generate an accurate 3D map of the environment are (i) acquirement of 3D information about the landmarks in the scene, and (ii) landmark identification and matching.

In Sect. 2, we review the current sensors utilized for 3D mapping. In Sect. 3, we outline the overall 3D sensory system architecture. Experimental results utilizing the proposed method are presented in Sect. 4.

2 Sensory systems for mapping USAR environments

In this section, we review the literature on current sensors that are being utilized to provide 3D range information as well as sensors that have been specifically developed to map rubble.

2.1 Current sensors for providing 3D range information

2.1.1 Stereovision

Among all the existing sensory techniques that can be potentially used for mapping, stereovision is probably the most studied method [14]. 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. Recently, stereo devices have been developed to provide real-time (at 30 fps) depth information, e.g. [36]. However, the real-time capabilities come at the cost of the accuracy of the stereo data [28]. In general, for 3D reconstruction, passive stereovision techniques depend heavily on cooperative surfaces, mainly on the presence of surface textures, and on ambient light [32]. With respect to USAR applications, [17] proposed a stereovision approach using two USB web cameras. The cameras provide a capture rate of 2.3–2.5 fps using a region-based approach for stereo matching. The system has been evaluated in a structured office environment and has yet to be applied to 3D scene mapping.

2.1.2 3D cameras

Recently, 3D cameras based on time-of-flight technologies have also been developed for the gathering of 3D data. The camera systems mainly consist of a modulated light source, in most cases infrared or near infrared [24] and a CMOS/CCD image sensor. The application of 3D cameras for USAR scenes is presented in [4,11,25]. For example, in [25], the CSEM Swiss Ranger SR-2 camera with a resolution of 124 × 160 pixels was used along with a 2-axis accelerometer (to measure pitch and roll) to develop a 3D mapping technique that was tested in a simulated collapsed house environment. The authors found that the sensor was not able to provide short range information and that a large error was present in the yaw rotation of the robot. The low resolution of the sensor also made it difficult to identify the majority of the objects in the scene. In [4] and [11], an additional 2D camera was used to provide more accurate texture information.

Newer versions of the Swiss Ranger have been developed with increased resolution; the Swiss Ranger SR-3000 [24] and the SR-4000 [18] both have a resolution of 176 × 144 pixels. However, in general, the pixel array size of these 3D cameras are limited and hence the resolution of both the 3D depth and 2D grayscale images is still relatively low as compared to other 3D sensors, where in particular (i) in the 3D images (sporadic) noise can be easily detected and can also increase as the distance from the camera to the scene increases, and (ii) nonlinear distortions caused by lens effects can be present, making it difficult to distinguish different objects in the scene. Sensor errors of 3D cameras are influenced by both the physical properties of the sensor and environmental conditions [29]. To effectively utilize 3D cameras in USAR environments, changes in the current hardware and software components of the system would be required to minimize image corruption and hence increase accuracy [3,28].

2.1.3 Laser scanners

Laser scanning technology 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 [30] or time-of-flight [35]. In [21], a 3D laser range finder consisting of a 2D laser scanner, a mount and a standard servo was built to map rescue environments. Multiple scans were taken by the laser range finder and merged to generate a 3D map of an environment. A single scan with a resolution of 361 × 176

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

data points took 4.5 s to obtain. In [26], an actuated laser range finder was used to perform 3D SLAM in the crashed parking lot at Disaster City, TX; namely, the laser scanner was rotated by a servo from −90° to +90° in intervals of 0.5° to provide a 3D point cloud containing approximately 195,000 points per sample. The time to acquire one full scan took approximately 32 s.

In general, the main disadvantage of using 3D laser scanning systems for robotic 3D mapping of USAR environments is that they are slow and 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 they cannot provide real-time range data acquisition. Furthermore, it has been found that it is difficult to perform reliable data association between current and previously detected features [12].

2.1.4 Sensors for mapping rubble

Due to the aforementioned limitations of stereovision, 3D cameras, and laser scanning for mapping in cluttered USAR environments, a handful of research projects have been proposed for the development of unique sensors for such environments. One such effort is by Kurisu et al. [13], who have proposed the use of two different laser range finders for 3D mapping of rubble. The first laser range finder consists of a ring of laser beam module and an omnivision CCD camera. The second sensor utilizes an infrared laser module with a CCD camera to capture the laser image and another camera for capturing 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) they rely on robot internal sensors to assist in mapping the USAR environment.

3 3D sensory system architecture

In this section, we will present the design of our proposed 3D structured light sensor within the context of our overall 3D sensory system architecture (Fig. 1). The sensor has the following advantages: (1) real-time capability—2D and 3D information of rubble filled environments can be acquired from the single sensor at a speed of 30 fps, (2) high-resolution mapping—with this sensor, a high-resolution 3D map can be built, and (3) invariant to ambient light illumination. In addition to the sensor, we will also present our novel landmark identification and matching approach. Lastly, we will discuss the visual simultaneous localization and mapping (SLAM) technique that was utilized to verify the application of our proposed system for 3D mapping.

3.1 Real-time 3D sensor

A major group of 3D sensing techniques is structured light, which includes various coding methods and employs varying number of coded patterns [7]. Unlike stereovision methods, structured light methods usually use processing algorithms that are much simpler. Furthermore, unlike its laser counterparts, capturing happens in a single step, allowing the entire surface to be digitized by a single acquisition. Structured light systems are commonly adopted because they are simple for recognition, sampling, modeling, and coordinate calculation.

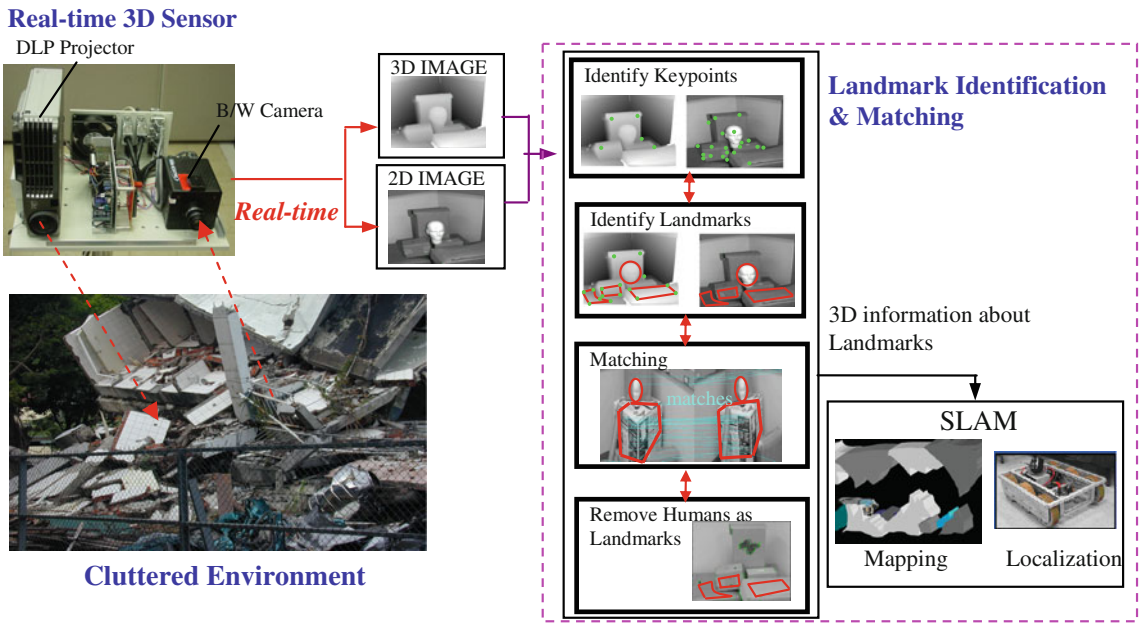

Fig. 1 Sensory system architecture

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

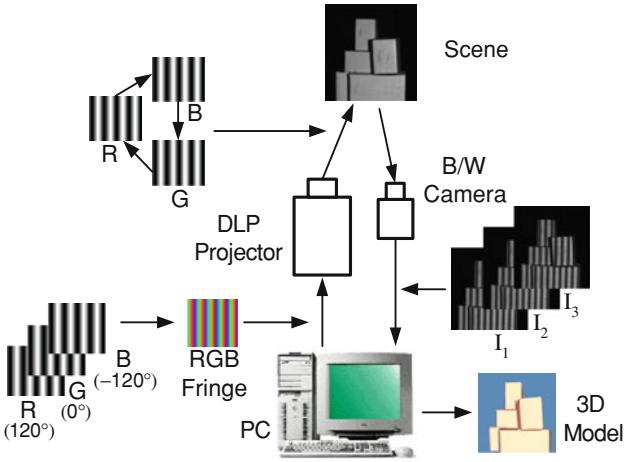

Fig. 2 System diagram

Therefore, it is more likely for them to achieve real-time performance (i.e., measurement and reconstruction). Many techniques in this group have been developed. Some use a single pattern, typically a color pattern, to boost speed. Examples include the color-encoded Moiré technique [31], rainbow 3D camera [6], color fringe projection method [9], and color structured light technique [38]. Even though these techniques show potentials for real-time 3D sensing, the results are affected to varying degrees by the variations of the object surface color. Other techniques 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 projection speed.

We have 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 [39,40]. The patterns are projected to the object with a switching speed of 360 fps. This sort of sensor is precisely what is needed to overcome the limitations of the aforementioned mapping sensors currently in use, but not much work has been done to apply this technology to the 3D mapping problem. Herein, we propose the development of a real-time 3D sensor based on the basic concept of the digital fringe projection and phase shifting technique for mapping of unknown USAR environments (Fig. 2). First, 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 color channels (R, G, B) in sequence repeatedly and rapidly over the scene. As a result, three grayscale fringe patterns (I_1, I_2, I_3) 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 and data modulation of an arbitrary point (x, y) in images captured with patterns (I_1, I_2, I_3) can be expressed with Eqs. (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. $I'(x, y)$ is used to generate the 2D intensity image. Once the phase information $\phi(x, y)$ is retrieved, the 3D information of the scene can be reconstructed by applying both a phase unwrapping algorithm and a triangulation algorithm. Meanwhile, the texture information of the landmarks can be easily retrieved from every three 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.

The DLP projector projects the fringe patterns at a frequency of 360 Hz. The B/W high-speed CCD camera synchronized with the DLP projector captures the fringe images at a lower frequency of 90 Hz, limited by the maximum frame rate of the camera. Each frame of the 3D scene is reconstructed and the texture image calculated using three consecutive fringe images. Therefore, the total 3D data and texture image acquisition speed of the sensor is 30 fps (the time to capture each 2D texture image is 25 ms). Together with the fast 3D reconstruction algorithms and parallel processing software we have developed, high-resolution real-time 3D measurement is realized at a frame rate of 30 fps and a data density of 532 × 500 points per frame. The range of the sensory system can be set by the focal length of the camera, the zoom of the projector, and the geometric relationship between the camera and the projector. Figure 3 presents the 3D information obtained by the sensor of a laboratory environment using the phase shifting method.

3.2 Identifying landmarks for mapping

Attempts have 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 scale invariant feature transform (SIFT) developed by Lowe [15]. This approach

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

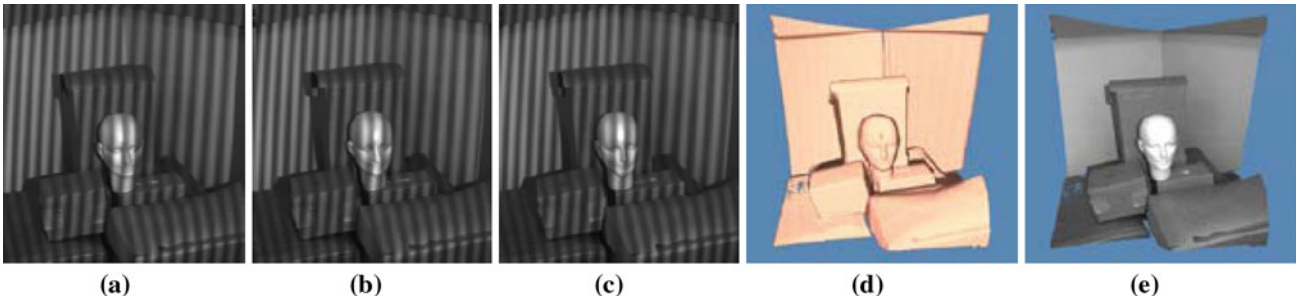

Fig. 3 **a–c** Phase-shifted fringe patterns, **d** 3D result, and **e** 3D result with B/W texture information

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. The SIFT approach consists of four main stages [16]: (i) scale-space extrema detection, (ii) keypoint localization, (iii) orientation assignment, and (iv) keypoint descriptor assignment. It is these keypoints that we will utilize to identify landmarks in our USAR environments. The SIFT 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. Se et al. [33] have proposed a vision-based SLAM method by tracking SIFT keys on 2D images and building a 3D map simultaneously utilizing a trinocular stereo system for an indoor environment, where the robot moves in an approximate 2D planar motion. Ellekilde et al. [4] have also proposed the use of SIFT-based 2D image matching in their work for data association within an EIF-based (extended information filter) SLAM framework for indoor search and rescue environments. Our own preliminary work, conducted by Nejat et al., in this area has included the development of a nearest neighbor 3D keypoint search method and also the utilization of Canny–Deriche edge detection for SIFT-based landmark identification [23,42,43].

In USAR environments, repetitive features may exist. As the robot moves, it must be able to determine whether different 3D sensor measurements correspond to the exact same feature in its environment; incorrect data association can induce extreme errors in mapping solutions. In our approach, we will utilize SIFT to identify landmarks in USAR environments. In the context of our work, landmarks are identified as objects or humans rather than the more traditional definition utilized in the literature which relates them merely to features of interest. Because visual systems are restricted to the sensing and processing of information which can be displayed as 2D projections, we also propose the utilization of high-resolution 3D depth images of the scene, provided by the proposed real-time 3D sensor, for landmark identification. This additional information is utilized to extract strong evidence in discontinuity between multiple objects detected in a scene to locate large distinguishable landmarks in a USAR environment for 3D mapping.

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 overall 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 first step of our landmark identification method consists of determining the keypoints of an image and their dimensional descriptors. In our proposed work, this will consist of two stages, finding the keypoints and descriptors for the 2D image and corresponding 3D depth image utilizing the SIFT algorithm. Both keypoints and descriptors are then stored for the two images. Figure 4 shows keypoints (green circles) that have been found on 2D and 3D images of a rubble-like environment with same size objects, and a large distinguishable object.
2. *Keypoint clustering* The clustering of keypoints is not only important in defining landmarks, but also in reducing the number of keypoints of interest. In general, due to shadowing effects and texture changes, a number of keypoints can be identified in the 2D images. Figure 4c shows the keypoints found on a box in an environment, with multiple keypoints on the flat surfaces of the box. In the 3D (i.e., depth) image, Fig. 4d, we can see that the keypoints on the flat surfaces are no longer present due to the fact 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 depth image in which for the latter image shadowing

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

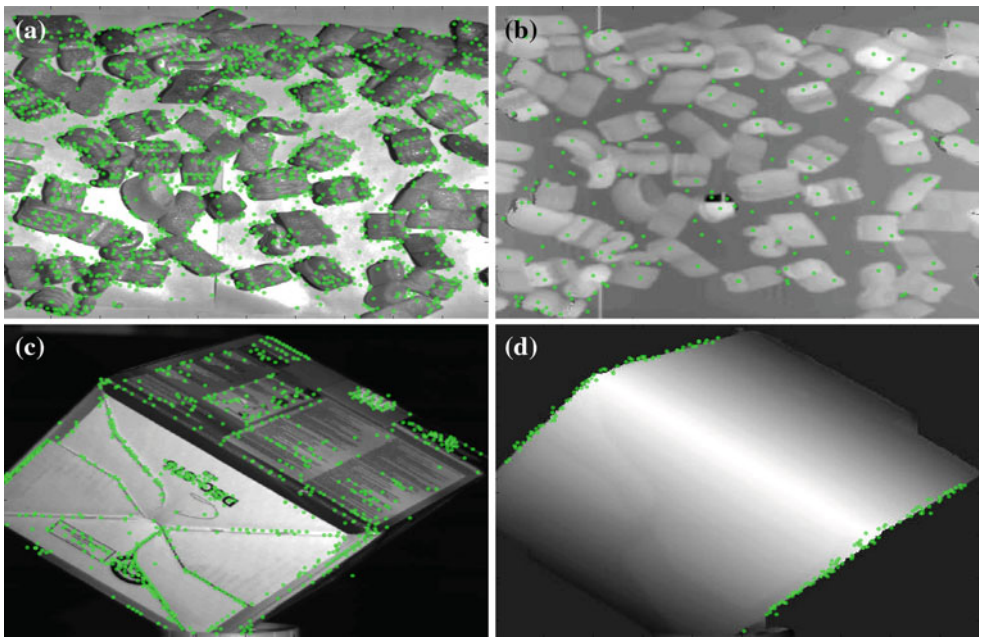

Fig. 4 **a** and **c** 2D image, and **b** and **d** 3D image of landmarks

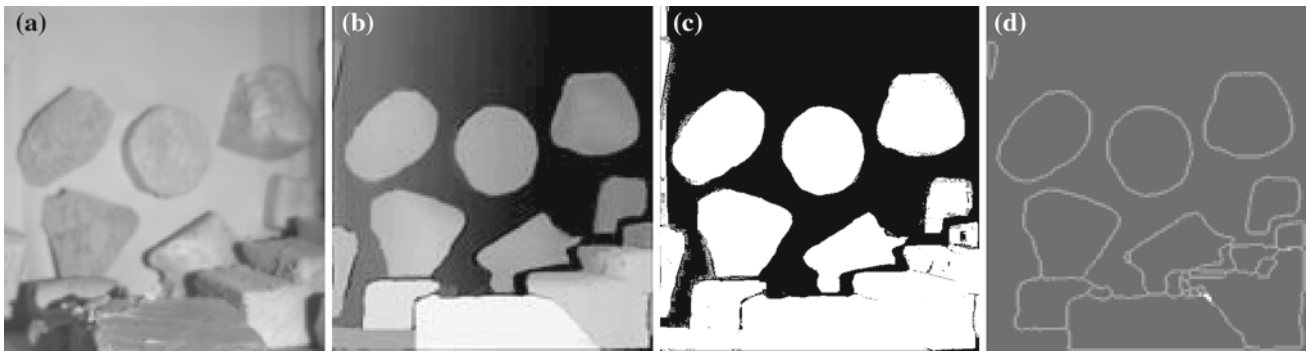

Fig. 5 **a** 2D image of the scene, **b** 3D image of the scene, **c** corresponding binary image, and **d** boundaries of objects

and texture effects are not present. 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 the 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 then 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.

Keypoints that are determined in the 3D image are grouped together based on the 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. We now illustrate our method utilizing the scene shown in Fig. 5. We initially store the following three parameters of each 3D keypoint: x location, y location, and depth in the matrix $\mathbf{A}_{ln}$, where l represents the number of keypoints and n represents the number of parameters (i.e. Table 1).

3.2.1 3D image analysis

Step 1 A binary-based background subtraction method is utilized to quickly and effectively determine and extract the boundaries of potential foreground objects in the scene to eliminate the unnecessary computation of keypoints associated with background noise (Fig. 5c). This background subtraction method separates foreground objects from background ones based on the depth information identified from the optimal range of the 3D sensor, however, it does not

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

Table 1 Keypoint parameter matrix **A**

Keypoint #	*x* position	*y* position	Depth
1	80.13	259.74	162
2	373.37	115.63	123
3	316.39	528.41	97
4	504.38	328.62	121
5	54.61	221.26	93
6	41.46	457.03	89
7	562.92	329.69	138
8	264.1	493.26	121
9	480.45	311.58	138

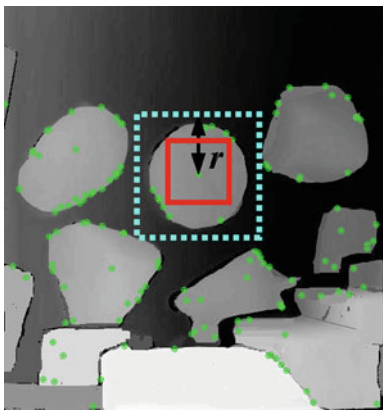

Fig. 6 Step 3 of algorithm: defining a search region

distinctly identify the edges of the foreground objects with similar depth information with respect to each other. Hence, we must implement further image conditioning techniques to identify distinguishable landmarks in the scene of interest.

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 [27]. Figure 5d shows the 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.

Step 3 Once the object boundaries are identified, our nearest neighbor search method is implemented. Random starting points,w_{ij} for the algorithm are chosen on the image. A square of side length $2r$ is drawn symmetrically around w_{ij} to search for its nearest neighbor keypoints (Fig. 6). If no keypoints are initially found, r is continuously incremented until keypoints are detected. Each detected keypoint and its parameters are stored in a temporary matrix B_{sn} for evaluation, where s represents the number of detected keypoints. There may be instances due to the location of these initial points on the image that the square may extend outside of the image boundaries, in which case only the portion of the square that encompasses the image is searched. The locations of several keypoints identified in Table 1 are presented in Fig. 7.

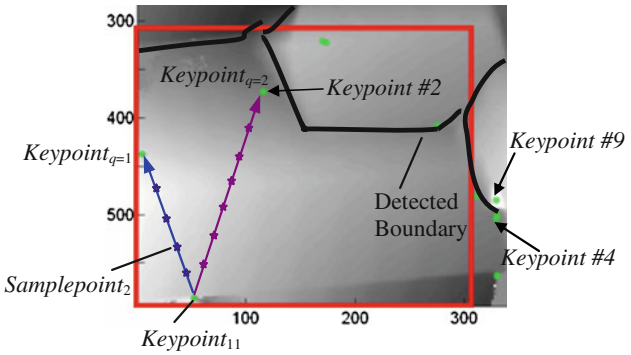

Fig. 7 Step 4 of algorithm: finding the nearest neighbor keypoint (keypoint #*i* represent the location of keypoints identified in Table 1)

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 $i = 1, 2, \ldots, N$ (Fig. 7). The nearest neighbor 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 $q = 1, \ldots, p-1, p$) and $keypoint_{11}$:

$$\text{Minimum\_depth\_value} = \min\left[\mathbf{A}(keypoint_{11}, 3), \mathbf{B}(keypoint_q, 3)\right], \tag{4}$$

$$\text{Maximum\_depth\_value} = \max[\mathbf{A}(keypoint_{11}, 3), \mathbf{B}(keypoint_q, 3)], \tag{5}$$

$$\text{Minimum\_depth\_value} - \text{threshold} < samplepoint_i\_depth < \text{Maximum\_depth\_value} + \text{threshold} \tag{6}$$

The objective of sampling multiple points between the keypoints is to ensure that boundaries of objects are not crossed. If a situation arises where a vector path from a keypoint to its nearest neighbor must cross an already existing vector, this latter vector must follow a different path. The most optimal path for clustering would be to follow along the edge, to provide the maximum surface area.

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 and this information is used along with

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

the sample points determined for the keypoint of interest in deciding whether the keypoint belongs to the cluster and its order within the cluster:

$$keypoint_{j(k+1)} = f[(samplepoint_i)_m, (samplepoint_i)_{k+1}, keypoint_{jm}], \quad (7)$$

where $m = 1, \ldots, k-1, k$. 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, and depth information stored from sample points. In order for a keypoint to be considered a part of 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 random starting points and Steps 3 and 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.

3.2.2 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. Because there is 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. 8). 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. Hence, also minimizing cluster computation time.

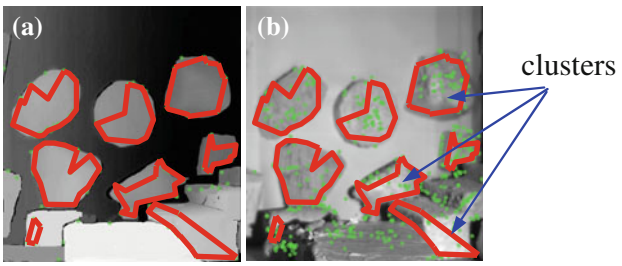

Fig. 8 Cluster results: **a** 3D image and **b** 2D image

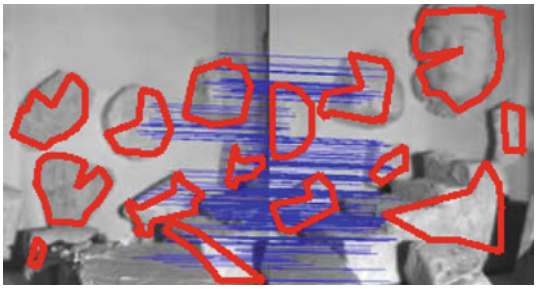

Fig. 9 Matching of clusters in different images

All clusters are identified as representing landmarks and are stored in a landmark database for landmark matching. The landmark database includes all 2D and 3D SIFT keypoint information corresponding to individual landmarks. This database is continuously updated when new landmarks are determined.

3. *Matching of clusters*. Matching of clusters relies on finding the same clusters in consecutive images by matching keypoints from the clusters from previous frames (as defined in our database) with ones in the new cluster of the current frame, we utilize the matching method proposed by Lowe [16], known as the Best-Bin-First (BBF) method. Herein, this can be achieved in terms of matching the key descriptors of the keypoints which can correspond to finding a set of nearest neighbors (NN) to a query point. The advantage of this method is its ability to handle high-dimensional spaces, i.e., the 128 dimensional descriptor vectors. Since individual SIFT keypoints are easily distinguishable, they can be matched correctly with an exception of a few false matches. These false matches can be further reduced in our method via the proposed clustering approach; namely, using the cluster association of the keypoints, we, in turn, can reduce false matches that may occur as a result of similar-looking features that are not associated with the same landmark. Figure 9 illustrates matching between clusters determined in two different viewpoints of a scene. The blue lines represent the keypoints that were matched within the two images. It is important to note in our work we are only interested in matching of the 2D keypoints identified within the clusters.

3.3 3D map building

The above procedure will provide identified 3D landmarks in a database for the robot to detect as it moves in the environment to create a 3D global map (Fig. 10). The most common approach in Visual SLAM is to create a database of single observed features for tracking. However, this approach

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

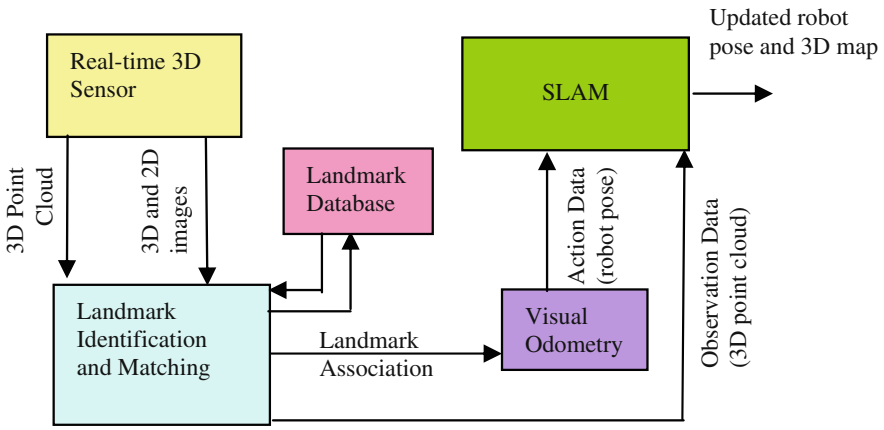

Fig. 10 Overview of mapping structure

can become very computationally expensive due to the large number of features that can be detected as the robot explores a large environment. Our unique approach eliminates the separate tracking of a large number of individual features by tracking small groups of associated features (i.e., landmarks) corresponding to foreground objects in the scene, and hence eliminates features due to background noise and initially unassociated features from the database.

To build the 3D map in world coordinates, the robot must be able to localize itself utilizing these determined landmarks. This can be achieved by stitching consecutive 3D range information corresponding to the landmarks provided by the sensor. Herein, we utilize the iterative closest point (ICP)—based SLAM method to align 3D range information provided by the 3D sensor at different locations to create a corresponding 3D map. The ICP-SLAM technique is a popular approach with many different variations and has been extensively used in the literature, i.e. [21,25,37]. It is important to note here that we choose to utilize ICP-SLAM due to its robustness in 3D environments and our ability to effectively illustrate our landmark-based approach, however, other SLAM methods could also be chosen and we will consider other approaches in our future work. Our proposed 3D sensing and landmark identification technique is not dependent on any single SLAM approach.

The ICP method proposed by Besl and McKay [1] is utilized to register all point clouds obtained by the robot as it moves in the environment into a common coordinate frame. In each iteration step, the ICP algorithm selects the closest points in two point sets as correspondences and calculates the transformation (R, t) for minimizing the equation [37]:

$$E(R,t) = \sum_{i=1}^{N_k} \sum_{j=1}^{N_l} w_{i,j} \left\| P_i - (RP_j + t) \right\|^2, \tag{8}$$

where N_k and N_l are the number of points in the point sets P_k and P_l, and $w_{i,j}$ are the weights for a point match. $w_{i,j} = 1$ when P_i is the closest point to P_j within a defined bound or $w_{i,j} = 0$ otherwise.

The approach we used, illustrated in Fig. 10, requires observation datasets from the sensor and the landmark identification and matching module, and action datasets for robot movement. The world coordinate frame is determined to be located at the robot's initial location in the scene.

Observation dataset

The observation dataset consists of the 3D range information of the scene provided by the 3D sensor via a point cloud with respect to the camera coordinate frame. The dataset consists of the complete 3D range information of the scene as represented by the point clouds. Within the overall dataset, 3D points corresponding to landmarks are marked as landmark range information. Due to the one-to-one correspondence between the 2D and 3D images, this information is readily found from the clusters identified on the 2D texture images. This dataset is utilized to generate an effective 3D map of the scene.

Action dataset

The action dataset contains the robot's estimated motion as determined using a visual odometry technique. In particular, the 3D coordinates of the same SIFT keypoints (SIFT pairs) representing the same landmarks in different images are utilized to estimate the 6 DOF ego-motion parameters (i.e., ΔX, ΔY, ΔZ, $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$) by using at least three pairs of SIFT keypoints. Since the position of the camera relative to the robot's coordinate frame is known, the transformation between two different locations of the robot is approximated. The approximated robot pose, herein, is used only as an input into the ICP algorithm which ultimately determines the robot's pose. To determine the robot pose solely via visual odometry is not accurate due to the fact that the

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

Fig. 11 Experimental set-up of the USAR scene

pose transformation is estimated only from a small number of SIFT pairs. However, it can provide the initial registration data needed for the ICP algorithm to successfully generate the required 3D map of the scene.

To reduce the computation time of ICP-SLAM, the 3D data points only associated with the landmarks in the scene, as defined in the observation dataset, are utilized to determine point correspondence within the ICP algorithm. However, the overall transformation is applied to the complete 3D range information to generate an overall map of the scene.

4 Experiments

A cluttered USAR-like scene (Fig. 11) was developed to verify the sensory system's robustness to generating 2D and 3D images of such a scene, and the feasibility of the proposed landmark identification method. The total area of the scene is defined to be approximately 4 m × 5 m. For the experiments, we utilized dim ambient lighting conditions.

The sensory system consisting of a DLP projector, in particular the PLUS U5-632 Digital projector with 1,024 × 768 resolution and 3,000 lumens light output and the Dalsa CA-D6-0512 B/W high-speed CCD camera (resolution 532 × 500) was placed on top of an all-terrain robot which navigated the scene (Fig. 12). The effective range of measurement of the system for these experiments was 0.7 ∼ 1.4 m with the current lens configuration of the camera and projector, which is sufficient for our USAR-like scene, since the area is cluttered and hence there is no long-range clear view into the rubble. The viewing angle for our feasibility experiments, presented herein, is defined to be approximately 16°. The sensory system provides high-resolution real-time 3D measurement at a frame rate of 30 fps. The rms error of the sensor was determined to be within 0.10–0.22 mm. The sensor was calibrated using a flat checkerboard placed at various positions and in various orientations. The size of each square on the checkerboard, which serves as the dimensional reference, is known precisely. By comparing the measured and true sizes of these squares located at various ranges, we can determine the measurement accuracy of the system in the entire measurement volume. For more details on the calibration of our sensory system, readers are referred to Huang and Han [8] and Zhang and Huang [41].

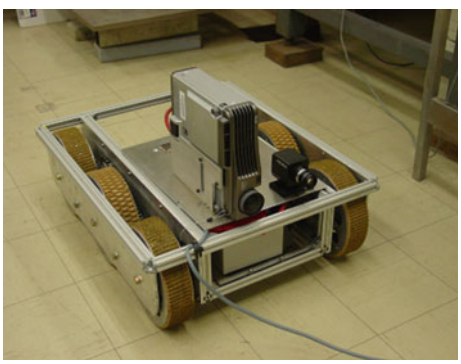

Fig. 12 The sensory system on a mobile robot

Based on the sensing principles utilized, the expected effective measurement range of the sensor can be increased to 3 m and the viewing angle can also be further increased by changing the hardware components of the sensor. This includes utilizing a wide angle and zoom lens for the camera and potentially replacing the projector lens with one that provides a good depth of focus within the range. The increase in the measurement range will result in an increase in the working distance of the system. This means that the view angle difference between the camera and the projector has to be reduced to maintain the same compactness of the system. This may potentially result in reduced resolution. However, this problem can be alleviated using a high-resolution camera. We believe that with the right optical design of the system, we should be able to obtain a measurement range of up to 3 m. We are currently in the process of optimizing our system to achieve this. This measurement range is greater than the range provided by the laser range finder used in [13] and consistent with the ranges provided by the 3D camera used in Ellekilde et al. [4] and Ohno et al. [25].

Since for the experiments presented herein the sensory system is taking consecutive images (at 30 fps) as the robot traverses 10–20 mm/s in the environment, there is an overlap of range data between the robot's different poses, allowing for the elimination of discontinuity in the data. While the

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

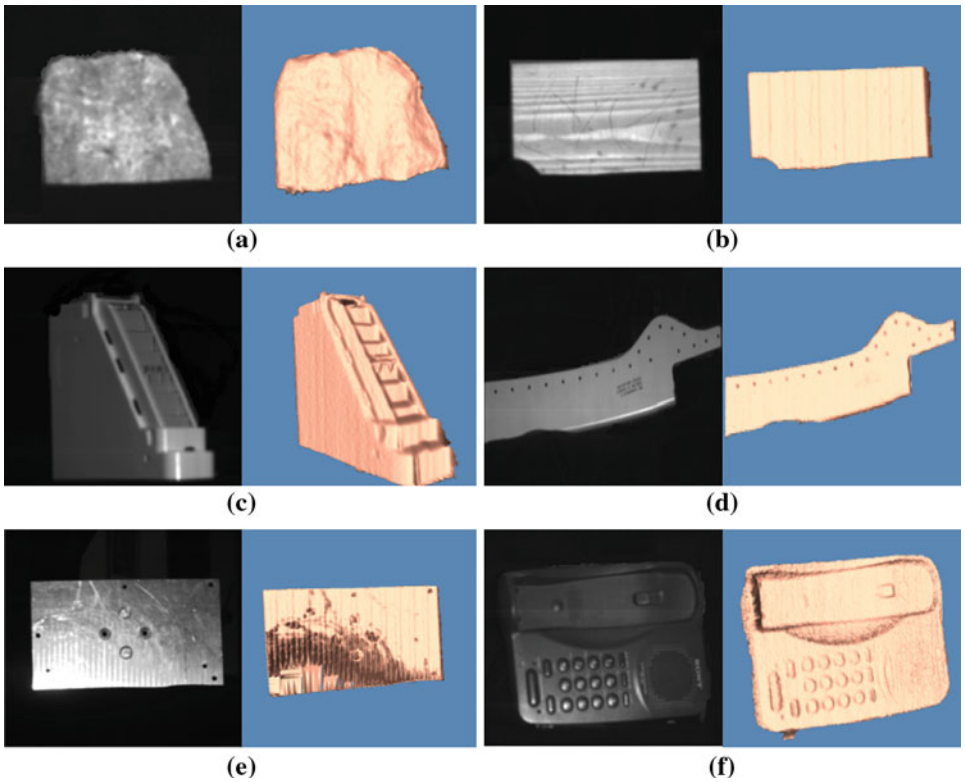

Fig. 13 2D and 3D images of different material properties: **a** concrete, **b** wood, **c** pink plastic, **d** matte metal, **e** shiny metal, and **f** black object

robot navigated through the environment, 2D and 3D images were taken in real-time. Utilizing the images taken by the sensor, the landmark identification and matching, and mapping algorithms were implemented.

4.1 3D sensory system

During the experiments, the sensory system's robustness to a USAR environment was assessed with regards to two main criteria: (i) object material composition, and (ii) scene illumination.

4.1.1 Object material composition

Our first set of experiments consisted of evaluating the sensory system's performance in the cluttered environment which consisted of different types of objects that may be found in a disaster scene. These objects were made from varying types of materials including wood, metal, plastic, brick, ceramic, concrete, paper, cardboard, plaster, rubber-like polymers and rocks. Furthermore, these objects ranged from bright to dark colors. As previously mentioned, the experiments were conducted in low lighting conditions. 2D images and 3D shape information generated by the sensor for a number of different types of materials present in the scene are shown in Fig. 13. In general, the sensory system can be utilized in scenes containing diffuse or matte or partially matte finish surfaces. However, the system similar to laser scanners is not very effective in imaging objects with specular surfaces as shown in Fig. 13e, where the black regions in the 3D model represent areas where 3D information could not be retrieved. Color does not have an effect on the proposed sensing method because the images taken of the scene are in grayscale. Furthermore, the phase shifting technique is utilized to separate texture information in the scene from the projected RGB pattern.

4.1.2 Scene illumination

Since the sensory system is an active system, the illumination conditions of the environment do not directly attribute to its sensing capabilities. In particular, the sensory system is capable of detecting objects within regular indoor lighting conditions, dim lit and dark environments. A set of experiments were implemented, herein, to demonstrate SIFT's performance utilizing the 2D images provided by the active sensory system at varying robotic poses within our low lighting USAR environment.

As the robot traverses the scene, its pose with respect to the objects in the scene is constantly changing, which may potentially cause a change in illumination as the 3D mapping sensor captures scene information. Experiments were performed to assess the performance of the SIFT method utilizing the sensory information captured during changes in robot pose. Previous studies utilizing SIFT during image illumina-

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

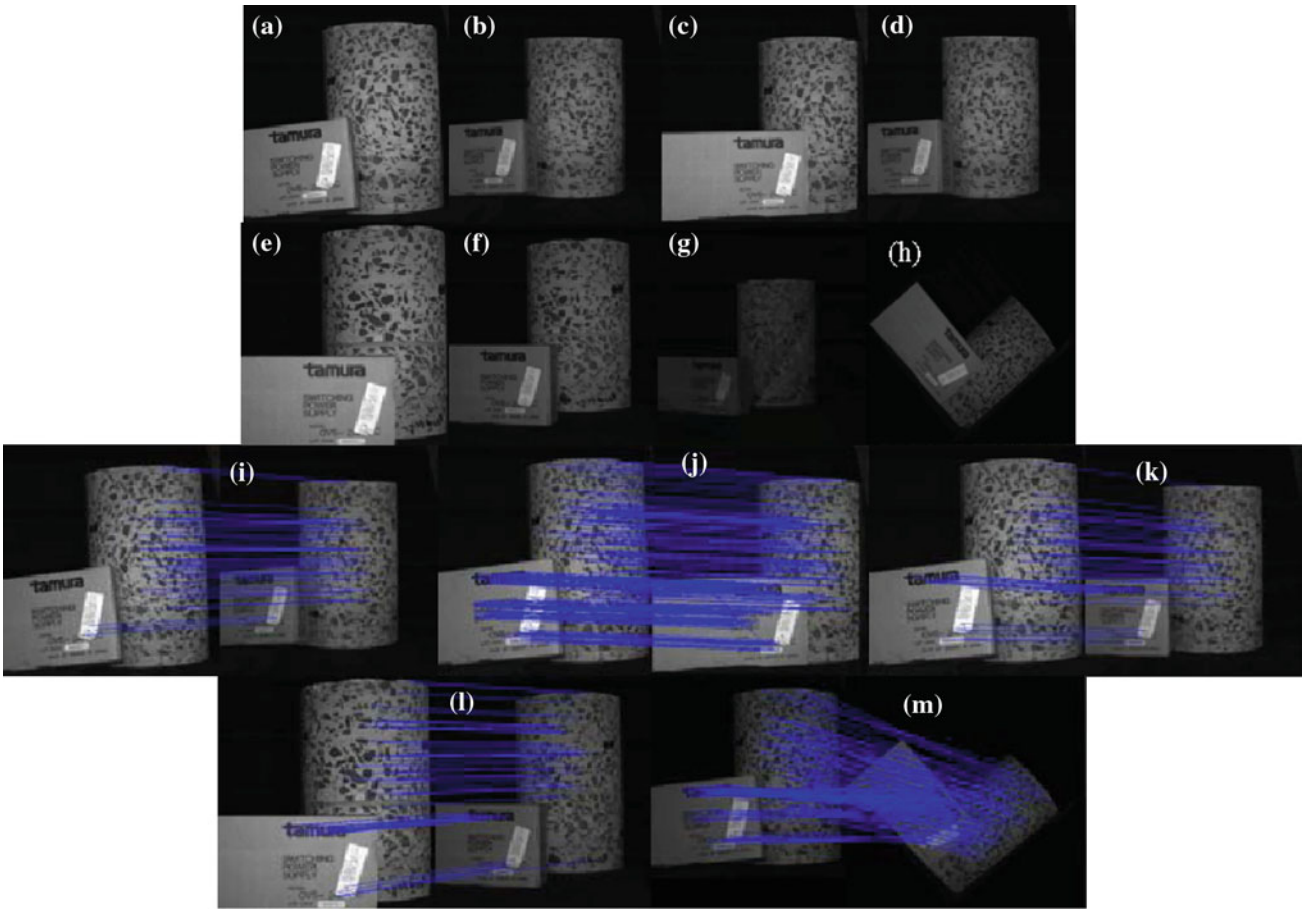

Fig. 14 **a–h** 2D images of objects in the USAR-like scene with varying scene illumination, and (i)–(m) SIFT matching results

tion changes have focused on physically changing the camera settings to generate illumination changes [19], however, illumination change in our scene is directly correlated to the pose of the robot during scene capturing. The relative pose changes for the robot were guided by scale and orientation parameters of the SIFT method and expected robot pose changes within the scene. In particular, we utilized a scene scale factor ranging from 0.2 to 2.7 and orientation changes $[\Delta x, \Delta y, \Delta z]$ within approximately [15, 50, 60] degrees with respective to the x, y, and z axes of the sensory system (as defined in Fig. 11).

Figure 14a–h presents 2D images taken at different robot poses with regards to two particular objects in the USAR-like scene. As can be seen in the images, the robot pose transformations introduced variations in illumination. SIFT identification and matching were performed on these different images to evaluate the performance of SIFT under such illumination changes (Fig. 14i–m). Our results found that we were able to successfully match objects in the scene effectively within a scale factor of 1.7 and orientation changes within approximately [15, 30, 50] degrees with respect to the three axes of the sensory system. As the scale factor and orientation changes both increase, the number of keypoint matches decreases, however, there still remain a significant number of matches. Beyond a scale factor of 1.7 and orientation changes especially in the y and z axes greater than 30° and 50° the number of matches is drastically reduced.

4.2 Identification and matching of landmarks for mapping

38 frames of 2D and 3D image pairs were taken in total by the sensory system as the robot traversed the scene and were utilized for landmark identification and matching. In general, 67 clusters were identified as landmarks in the scene, for which 53 were matched successfully in different frames. For these experiments, the clusters that had more than three correct keypoint matches were recognized by the algorithm to be the same landmark in the scene. Images from two poses of the robot are shown herein to illustrate the effectiveness of the proposed landmark identification and matching methodology (Fig. 15). An an average of 184 and 964 keypoints was determined in the 3D and 2D images, respectively. Six clusters were identified at each pose. All three (#1,3,5) clusters representing the same landmarks at both poses were matched

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

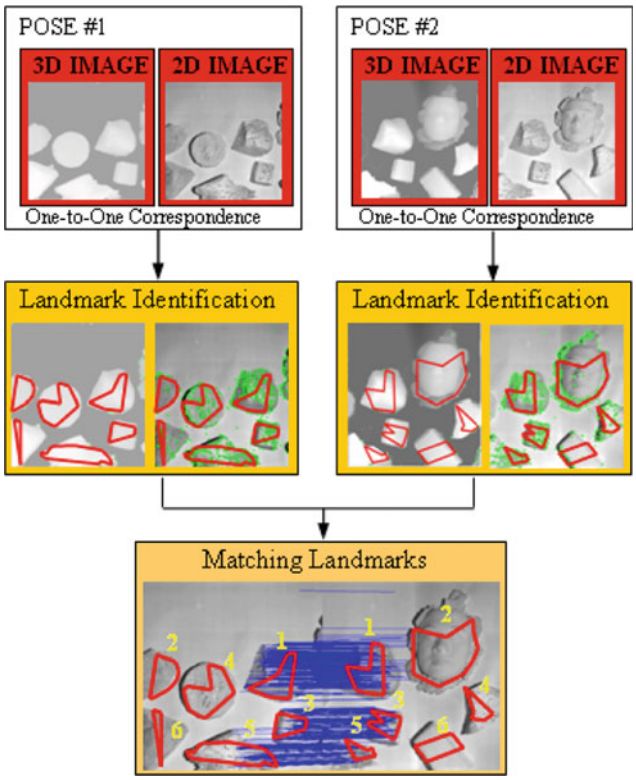

Fig. 15 3D and 2D images of two poses, and landmark identification and matching

effectively between the two poses. Figure 16 presents the landmark identification and matching results for ten other robot poses in the scene. We have also determined the recall rate of the clusters in the scene to be 85%. Herein, the recall rate represents the ratio of matched clusters between two images from the scene and the total number of clusters identified in the overlapped region of these two images. We then compared this recall rate to that determined by directly matching individual SIFT keypoints (without the use of clusters) in the overlapped region of two images, and found the recall rate for the matching of individual keypoints in the scene to be only 42%.

All 38 frames were used in the ICP-SLAM algorithm. In general, the number of 3D points generated by the sensory system for each frame is defined to be 266,000 (532 × 500). For efficiency, we only utilize 3D points associated with landmark clusters for data registration, hence, on average utilizing 122,892 3D data points per frame. Figure 17a presents the generated final map of the scene including zoom in views of 2D and 3D information generated by the sensory system at various locations along the robot's path. The sensor provides high-resolution sensory information as is evident by the details of the landmarks in the 2D and 3D results presented in the zoom in views. We have also measured the location of the boards providing the real boundary of the scene (as seen in Fig. 11) and superimposed their approximate locations onto the reconstructed 3D map as can be seen in Fig. 17b. The results show that the 3D map is consistent with the real USAR-like scene. From ICP-SLAM, we have determined the maximum accumulated error for the robot's location to be 4.8% of the robot's total travel distance of 941 cm.

While the 2D and 3D sensory information was obtained at 30 fps, the total processing time for landmark identification and mapping took 807 s of CPU time on a Pentium IV 3.0 GHz 1.0G RAM system utilizing an MATLAB platform. The breakdown of this time is presented in Table 2. The average processing time for landmark identification for a single image was determined to be approximately 15 s and the average processing time for landmark matching between two consecutive images once these landmarks have been identified was determined to be approximately 2.7 s. The overall computation time is more efficient as compared to the ICP-SLAM work conducted with similar size 3D point data sets obtained by laser scanners [22]. This time can be further reduced by the optimization of the proposed algorithms and their implementation in a C++ platform environment which is a part of our future work.

4.3 Performance comparison of landmark identification technique

To validate the use of our proposed landmark identification method, a performance comparison study has been conducted between our proposed clustering-based landmark identification approach versus a more traditional non-clustering-based approach. In particular, the non-clustering approach consists of the following steps: (i) 2D SIFT keypoint identification and matching; (ii) RANSAC (RANdom SAmple Consensus) to identify inliers, and (iii) ICP-based SLAM. RANSAC is an iterative method used to estimate parameters of a mathematical model from a set of observed data which contains outliers [5]. RANSAC has been extensively used in identifying SIFT inliers. Our comparison experiments consisted of implementing our proposed clustering-based approach and the non-clustering based approach for two different scenarios: (i) when the percentage of inlier features is 91%, and (ii) when the percentage of inlier features is 22% (Table 3). The computation time for our proposed method including keypoint identification, landmark identification and ICP-SLAM algorithm was determined to be approximately 807 s regardless of the defined percentage of inliers. For the non-clustering method and ICP-SLAM algorithms, in which 91% of feature points were inliers, the implementation time was 588 s for 1,001 iterations. However, since RANSAC is an iterative-based method, this time can greatly vary and was increased to 1,090 s for 5,862 iterations for the experiments where 22% of inlier features were present. The main disadvantage

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

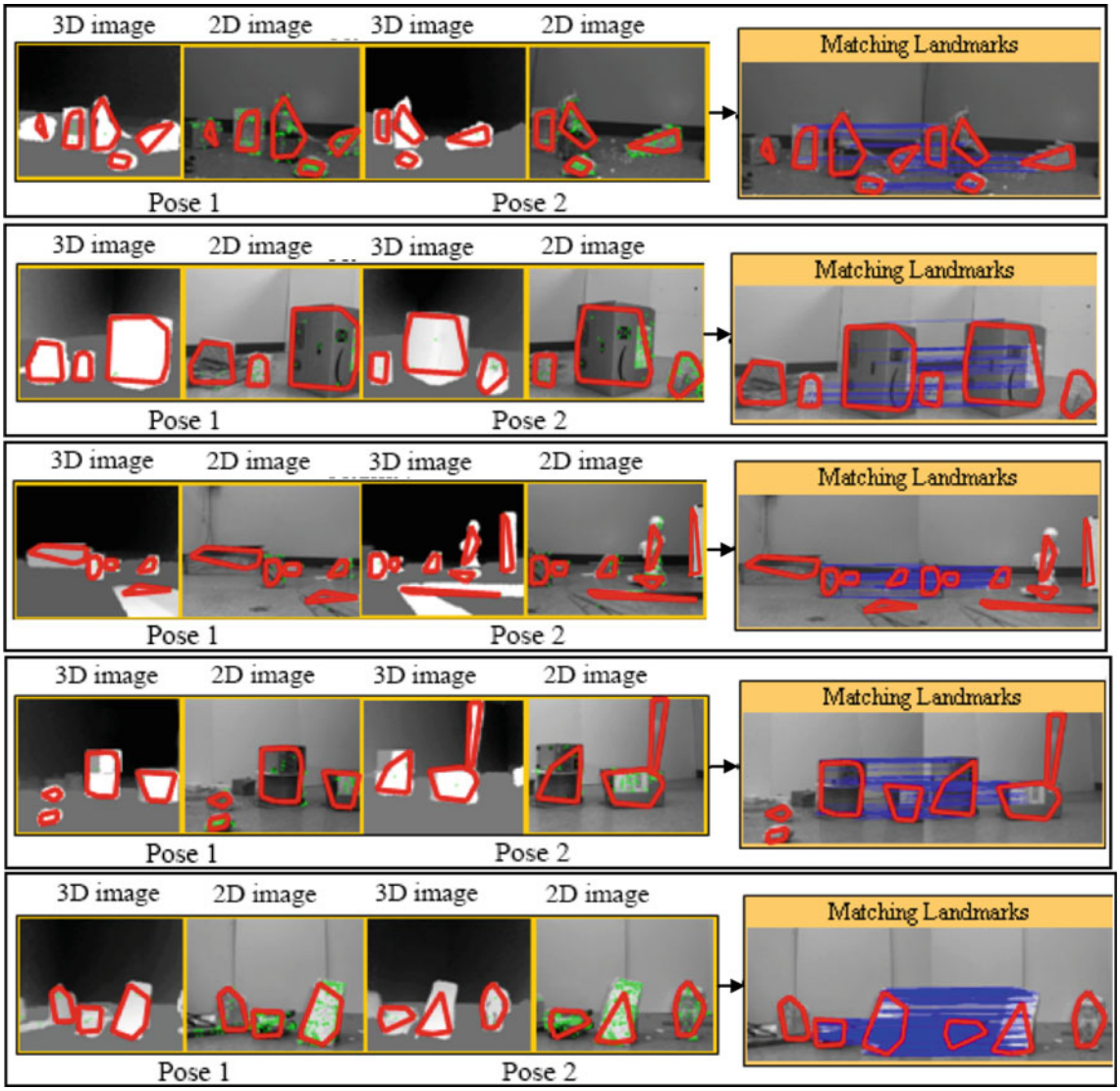

Fig. 16 Landmark identification and matching results for other poses in the scene

of RANSAC is its iterative inherency, where computation time can greatly increase in varying situations, in particular when the percentage of inliers is reduced. Even though an upper bound on time can be applied to RANSAC, the solution obtained at this upper bound may not be optimal. Our geometric constraint-based clustering approach for landmark identification is effective even when smaller percentages of inliers are present.

Furthermore, it is important to note that even though the clustering stages on their own do add computation time, they however, assist in decreasing the computation time of other required steps such as 2D SIFT keypoint matching and ICP-SLAM. Lowe [16] found that a typical 2D image can contain 2,000 or more features, which are a result of different objects in the scene and background noise. Since our approach only utilizes a fraction of the 2D keypoints, namely by eliminating the keypoints due to background noise and only utilizing a portion of the keypoints on the foreground objects, we are able to minimize the time it takes to cluster and match 2D keypoints. The utilization of such a small number of keypoints for matching is justified in Lowe [16] where it has been found that reliable object recognition is possible with as few as three features. Several experiments conducted in our work have shown that on an average, we can reduce the number of 2D keypoints utilized for matching by 41% versus all the detected keypoints when only considering the 2D keypoints in the defined clusters on the foreground objects. This reduction in keypoints further reduces matching computation time by approximately 42%. Furthermore, when comparing to matching of all keypoints detected on just the foreground objects versus those defined within our clusters, this reduction can be up to 40%. In addition, the number of 3D data points utilized by the ICP-SLAM algorithm can also be reduced by up to 46%, due to the fact that we eliminate the

This is the author's version of an article that has been published in Intel Serv Robotics (2011) 4:119–134 DOI 10.1007/s11370-010-0082-3.
Changes were made to this version by the publisher prior to publication.
The final version of the paper is available at https://link.springer.com/article/10.1007/s11370-010-0082-3

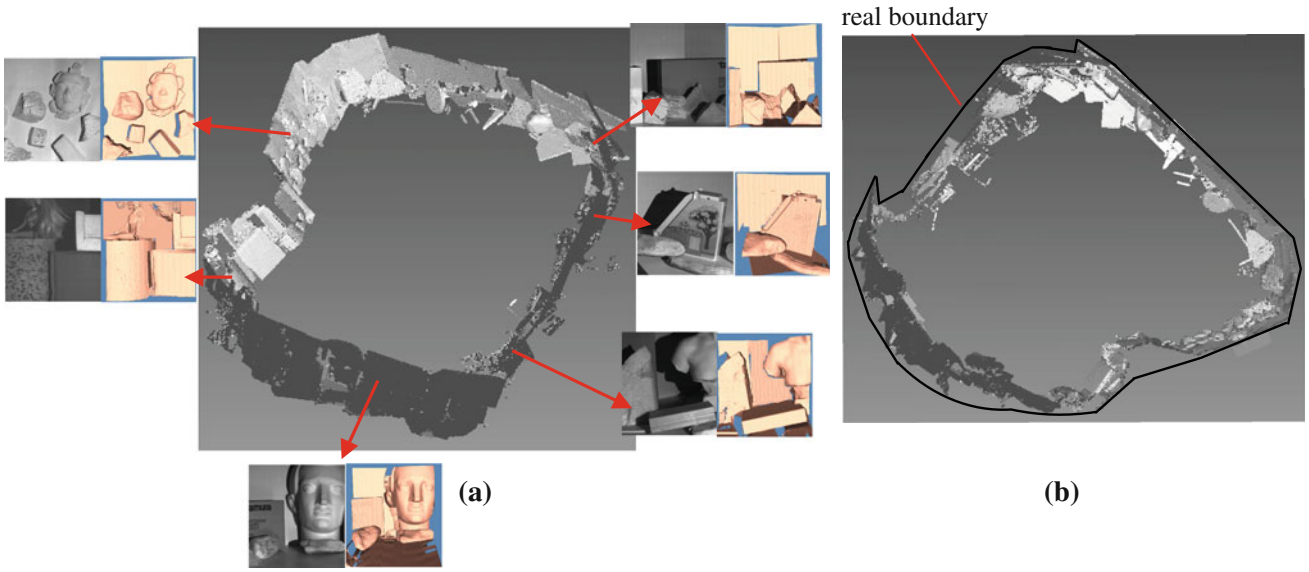

Fig. 17 Two different views of the generated 3D map of the scene including **a** zoom in views of sensory information in different regions of the map, and **b** superimposed boundary of the real scene

Table 2 Processing time for landmark identification and mapping

Stages of method	Time
Keypoint identification	
2D image	132 s
3D image	146 s
3D image analysis	
Step 1	169 s
Step 2	110 s
Steps 3 and 4	19 s
2D image analysis	102 s
ICP SLAM	129 s
Total	807 s

Table 3 Computation times for clustering and non-clustering techniques

Percentage of Inlier features (%)	Computation time (s)	
	Our proposed clustering approach	Non-clustering approach
91	807	588
22	807	1,090

use of background points during the data registration of this stage, which can, therefore, also reduce the SLAM algorithm computation time.

5 Conclusions

In this paper, we propose the first application of using a structured light sensor for sequential map building in cluttered USAR environments. The sensor can provide real-time (at a speed of 30 fps) high-resolution 2D and 3D sensory information and can work effectively in dim and dark environments. We have also developed a unique SIFT-based clustering method to analyze the 2D and 3D images taken by the sensor for effectively identifying distinguishable landmarks in a scene and matching the landmarks for 3D mapping. The preliminary experiments presented show the potential of the proposed sensory system for cluttered unknown environments. The sensory system was effective in providing the sensory information needed to build a 3D map of a USAR-like scene. Our future work will consist of optimizing both the hardware and software aspects of the proposed sensory system to implement the system in larger more complex environments to verify its performance in these types of environments.

References

1. Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14:239–256
2. Bosse M, Newman P, Leonard J, Soika M, Feiten W, Teller S (2003) An Atlas framework for scalable mapping. In: IEEE international conference on robotics and automation (ICRA), pp 1899–1906
3. Craighead J, Day B, Murphy R (2006) Evaluation of Canesta's range sensor technology for urban search and rescue and robot navigation. Technical Report, CRASAR-TR2006-2, pp 1–5
4. Ellekilde L-P, Huang S, Miró JV, Dissanayake G (2007) Dense 3D map construction for indoor search and rescue. J Field Robot 24(1):71–89
5. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24:381–395

6. Geng ZJ (1996) Rainbow 3-D camera: new concept of high-speed three vision system. Opt Eng 35:376–383
7. Harding KG (1988) Color encoded Moiré contouring. In: Optics, illumination, and image sensing for machine vision III, Proceedings of SPIE, vol 1005, pp 169–178
8. Huang PS, Han X (2006) On improving the accuracy of structured light systems. Proceedings of SPIE two- and three-dimensional methods for inspection and metrology IV, optics east, 6382
9. Huang P, Hu Q, Jin F, Chiang FP (1999) Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. Opt Eng 38:1065–1071
10. Ishida H, Nagatani K, Tanaska Y (2004) Three-dimensional localization and mapping for a Crawler-type mobile robot in an occluded area using the scan matching method. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 449–454
11. Kadous M, Sammut C, Sheh R (2005) Caster: a robot for urban search and rescue. Australas Conf Robot Autom 1–10
12. Karlsson N, Di Bernardo E, Ostrowski J, Goncalves L, Pirjanian P, Munich ME (2005) The vSLAM algorithm for robust localization and mapping. In: IEEE international conference on robotics and automation, (ICRA), pp 24–29
13. Kurisu M, Yokokohji Y, Oosato Y (2005) Development of a laser range finder for 3D map-building in rubble the 2nd report: development of the 2nd prototype. In: IEEE international conference on mechatronics and automation, pp 1842–1847
14. Lobo J, Queiroz C, Dias J (2003) World feature detection and mapping using stereovision and inertial sensors. J Robot Auton Syst 44:69–81
15. Lowe DG (1999) Object recognition from local scale-invariant features. In: International conference on computer vision, pp 1150–1157
16. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
17. McKinnon B, Baltes J, Anderson J (2005) "A region-based approach to stereo matching in USAR". In: Proceedings of RoboCup-05, Osaka
18. MESA Imaging (2009) SR4000 Data Sheet Rev. 3.0. Available at http://www.mesa-imaging.ch/dlm.php?fname=pdf/SR4000_Data_Sheet.pdf
19. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 27(10):1615–1630
20. Murphy RR (2004) Human–robot interaction in rescue robotics. IEEE Trans Syst Man Cybern Part C Appl Rev 34(2):138–153
21. Nuchter A et al (2005a) Mapping of rescue environments with Kurt3D. IEEE Int Workshop Rescue Robot (SSRR '05):158–163
22. Nüchter A, Lingemann K, Hertzberg J, Surmann H (2005b) 6D SLAM with approximate data association. In: International conference on advanced robotics (ICAR), pp 242–249
23. Nejat G, Zhang Z (2006) Finding disaster victims: robot-assisted 3D mapping of urban search and rescue environments via landmark identification. In: IEEE the 9th international conference on control, automation, robotics and vision (ICARCV), pp 1381–1386
24. Oggier T, Büttgen B, Lustenberger F, Becker G, Rüegg B, Hodac A (2006) SwissRanger SR3000 and first experiences based on miniaturized 3D-TOF cameras. Available at http://www.mesa-imaging.ch/pdf/Application_SR3000_v1_1.pdf
25. Ohno K, Nomura T, Tadokoro S (2006) Real-time robot trajectory estimation and 3-D map construction using 3-D camera. In: IEEE/RSJ international conference on intelligent robots and systems, pp 5279–5285
26. Pathak K, Birk A, Vaskevicius N, Pfingsthorn M, Schwertfeger S, Poppinga J (2010) Online 3D SLAM by registration of large planar surface segments and closed form pose-graph relaxation. J Field Robot 27:52–84
27. Pavillon N (2003) Edge detector. Available at http://bigwww.epfl.ch/demo/subpixeledge/start.php
28. Poppinga J, Birk A, Pathak K (2010) A characterization of 3D sensors for response robots. Robocup 2009 (Lect Notes Artif Intell 5949):264–275
29. Rapp H (2007) Experimental and theoretical investigation of correlating TOF-camera systems. Faculty of Physics and Astronomy, University of Heidelberg, Germany
30. Rioux M (1984) Laser range finder based on synchronized scanners. J Appl Opt 23(21):3837–3844
31. Salvi J, Pages J, Batlle J (2004) Pattern codification strategies in structured light systems. Pattern Recognit 37(4):827–849
32. Se S, Jasiobedzki P (2006) Photo-realistic 3D model reconstruction. In: Proceedings of IEEE international conference on robotics and automation (ICRA), pp 3076–3082
33. Se S, Lowe D, Little J (2001) Vision-based mobile robot localization and mapping using scale-invariant features. In: IEEE international conference on robotics and automation (ICRA), pp 2051–2058
34. Smith R, Self M, Cheeseman P (1990) Estimating uncertain spatial relationships in robotics. Auton Robot Vehicles 167–193
35. Tripp J, Ulitsky A, Pashin S, Mak N, Hahn J (2003) Development of a compact, high-resolution 3D laser range imaging system. Proc SPIE Int Soc Opt Eng 5088:112–122
36. Videre Design LLC (2010) Apparen family of stereo vision cameras and software. Menlo Park, CA. Available at http://www.videredesign.com/
37. Wulf O, Nüchter A, Hertzberg J, Wagner B (2008) Benchmarking urban six-degree-of-freedom simultaneous localization and mapping. J Field Robot 25(3):148–163
38. Zhang L, Curless B, Seitz SM (2002) Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: The 1st IEEE international symposium on 3D data processing, visualization, and transmission, pp 24–36
39. Zhang S, Huang P (2004) High-resolution, real-time 3D shape acquisition. In: IEEE computer vision and pattern recognition workshop on realtime 3D sensors and their uses, pp 28–37
40. Zhang S, Huang PS (2006a) High-resolution, real-time 3-D shape measurement. Opt Eng 45(12):123601-1–123601-8
41. Zhang S, Huang PS (2006) Novel method for structured light system calibration. Opt Eng 45(8):083601
42. Zhang Z, Guo H, Nejat G, Huang P (2007) Finding disaster victims: a sensory system for robot assisted 3D mapping of urban search and rescue environments. In: IEEE international conference on robotics and automation (ICRA), Rome, Italy, pp 3889–3894
43. Zhang Z, Nejat G (2008) Robot-assisted intelligent 3D mapping of unknown cluttered search and rescue environments. In: IEEE/RSJ international conference on intelligent robots and systems, pp 2115–2120