Robotica: (2020) volume 38, pp. 350–373. © Cambridge University Press 2019
doi:10.1017/S0263574719000675

Three-Dimensional Reconstruction Based on Visual SLAM of Mobile Robot in Search and Rescue Disaster Scenarios

Hongling Wang†‡, Chengjin Zhang¶*, Yong Song¶*, Bao Pang† and Guangyuan Zhang‡

†*School of Control Science and Engineering, Shandong University, Ji'nan 250101, China*
E-mail: pang_bao11@163.com
‡*School of Information Science and Electronic Engineering, Shandong Jiao Tong University, Ji'nan 250357, China. E-mails: wanghongling@sdjtu.edu.cn, zhangguangyuan@sdjtu.edu.cn*
¶*School of Mechanical, Electrical and Information Engineering, Shandong University at Weihai, Weihai 264209, China*

(Accepted April 10, 2019. First published online: May 21, 2019)

SUMMARY
Conventional simultaneous localization and mapping (SLAM) has concentrated on two-dimensional (2D) map building. To adapt it to urgent search and rescue (SAR) environments, it is necessary to combine the fast and simple global 2D SLAM and three-dimensional (3D) objects of interest (OOIs) local sub-maps. The main novelty of the present work is a method for 3D OOI reconstruction based on a 2D map, thereby retaining the fast performances of the latter. A theory is established that is adapted to a SAR environment, including the object identification, exploration area coverage (AC), and loop closure detection of revisited spots. Proposed for the first is image optical flow calculation with a 2D/3D fusion method and RGB-D (red, green, blue + depth) transformation based on Joblove–Greenberg mathematics and OpenCV processing. The mathematical theories of optical flow calculation and wavelet transformation are used for the first time to solve the robotic SAR SLAM problem. The present contributions indicate two aspects: (i) mobile robots depend on planar distance estimation to build 2D maps quickly and to provide SAR exploration AC; (ii) 3D OOIs are reconstructed using the proposed innovative methods of RGB-D iterative closest points (RGB-ICPs) and 2D/3D principle of wavelet transformation. Different mobile robots are used to conduct indoor and outdoor SAR SLAM. Both the SLAM and the SAR OOIs detection are implemented by simulations and ground-truth experiments, which provide strong evidence for the proposed 2D/3D reconstruction SAR SLAM approaches adapted to post-disaster environments.

KEYWORDS: SLAM; 3D reconstruction; RGB-D visual system; SAR environments; Robot mapping.

1. Introduction

Simultaneous localization and mapping (SLAM) involves a mobile robot generating a map of its surrounding environment and updating it online while simultaneously using the map to locate itself in the uncertain environment. Lin and Wang[1] demonstrated a speeded-up robust features (SURF) SLAM approach for search and rescue (SAR) scenarios, and the provision of three-dimensional (3D) images to a teleoperator via a mobile robot has been established for SAR sits;[2] indeed, 3-D information

* Corresponding authors. E-mails: cjzhang@sdu.edu.cn, songyong@sdu.edu.cn

is a key feature for understanding robot exploration.[3] The Center for Robot-assisted Search and Rescue responded within 6 h to the World Trade Center disaster in September 2001; this was the first known use of robots for urban SAR (USAR). The use of hierarchical analysis in visual SLAM research stems from four Japanese researchers who study robotics for national earthquake SAR post-disaster relief planning.[4] Most work in SLAM has focused on map building, including generating two-dimensional (2D) or topological graphical representations. Tracked or wheeled crawler mobile robots have become the best tools for maneuvering in the rubble of a collapsed mine; these are more rugged, permit general payload exchanges, and incorporate autonomous exploration, localization, and mapping capabilities.[5]

SLAM is useful for mobile robot exploration, mapping and updating in unstructured SAR collapsed environments, and simultaneously collecting and storing information about the robot's location. Laser-based representative 2D SLAM algorithms are currently available in robot operation systems (ROSs) and include HectorSLAM, Gmapping, KartoSLAM, CoreSLAM, and LagoSLAM.[6] In SAR scenarios, a complete system was a popular approach for acquiring textured 3D point clouds using a laser range finder (LRF) and a charge coupled device (CCD) camera.[7] A SAR object is described by the characteristics of (i) color, (ii) texture, (iii) shape, and (iv) size, and these four parameters represent the object's essential features. To explore a SAR scene, a mobile robot requires 3D map of the rubble for the SLAM process. In ref. [4], two wireless CCD cameras were used to configure a pair of stereo cameras, and a 3D rubble map was constructed by teleoperated mobile robots. The LRF mounted on a mobile robot generates the range data surrounding the robot, and the iterative closest point (ICP) algorithm uses two consecutive frames to estimate the motion of the camera.[8] The visual SLAM (V-SLAM) algorithm uses a monocular framework to stabilize the mobile robot's exploration in six degrees of freedom (6-DOFs), which overcomes the problem of both the drift and the Global Position System (GPS) dependency.[9] Nagatani et al.[10] proposed the sphere digital elevation map as a model to represent environmental information and to allow the robot to localize itself in the resultant map, and a system for mapping and virtual reconstruction was developed as part of a robot for USAR.[11]

A problem with 3D reconstruction is often a lack of robustness when building detailed 3D dense maps. RGB-D (red, green, blue + depth) mapping can provide a framework for dense environment 3D maps despite the limited depth precision and field of view. The improved ICP (IICP) algorithm takes advantage of the rich information contained in RGB-D data; it integrates vision and color information in depth-based techniques to produce robust frame matching and loop-closure detection. The robot can use an RGB-D camera to generate extra information to store in the map. During robot SLAM, this depth information can be added to identify objects of interest (OOIs) such as victims, gates, and stairs in a SAR scenario. Consequently, as a robot builds a map while exploring an area, the positions of the OOIs and the robot itself are simultaneously located accurately.

In the present study, we proposed two novel processing methods, namely fusion-3D and the 3D reconstruction. Fusion-3D uses the 2D image, a 3D point cloud, and texture information to extract 3D image features; while 3D reconstruction uses the point cloud data to generate a texturized mesh form. In these systems, the 3D point clouds are pre-processed, and the mesh is simplified to fuse into a consistent topological 2D/3D map. Consequently, two types of map, namely dense and sparse, are produced: dense maps are generated from 3D point clouds, and sparse maps are generated primarily from features extracted from images. A fast simple global 2D SLAM is conducted in a SAR environment, and 3D OOI local sub-maps are reconstructed. The 2D map planar is segmented into local sub-maps, and 3D point clouds are pre-processed to a consistent topological 2D map. Based on the 2D map, 3D-object sparse features are extracted and reconstructed, thereby allowing 3D SAR OOIs and loop-closure detection to be identified.

The rest of this paper is organized as follows. In Section 2, the main methods for 3D reconstruction are elaborated. In Section 3, visual reconstruction algorithms are derived. In Section 4, simulations of 3D mapping reconstruction are conducted using a virtual mobile robot on the MobileRobots platform. In Section 5, experiments are implemented for 3D reconstruction SLAM with a crawler robot, Pioneer LX, and the Amigo mobile robot. In Section 6, the results of the simulations and experiments are discussed. Finally, conclusions are drawn and future work is previewed in Section 7.

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

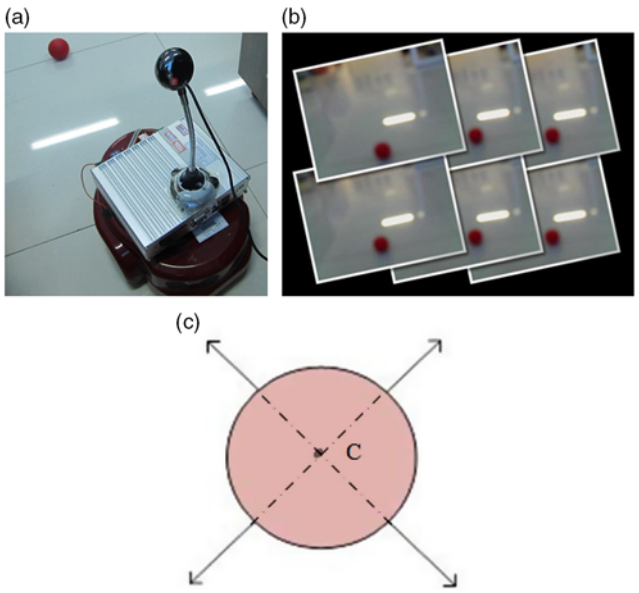

Fig. 1. Optical flow representation of depth translational motion. (a) Amigo mobile robot equipped with camera. (b) Consecutive image frames of a red ball. (c) Optical flow changes represent the depth and translational movement.

2. Three-Dimensional Reconstruction Methods in Unstructured Environments

2.1. Optical flow approaches in visual SLAM

Reconstructing a 3D-object local map from online visual images is an attractive technique in SLAM. We have developed a 3D object reconstruction method that represents robot motion by a series of consecutive image frames. Optical flow reflects the changes in the images due to motion during a time interval, and the optical flow field is the velocity field that describes the 3D motion of objects across consecutive 2D image frames. A valuable application of optical flow computation is in the analysis of robot movement relative to SAR objects in a post-disaster environment. A mobile-robot vision system is seen as a moving observer, and optical flow computation can be used to determine the current distance (i.e., depth information) of an OOI from the robot's position. Motion estimation relies on the brightness constancy principle assumed in optical flow techniques; it handles the appearance variation caused by illumination and pose change using parametric models, which extends a gradient-based optical flow algorithm.

We first developed the application of optical flow theory to SAR SLAM in post-disaster environments, and Algorithm 1 is the proposed procedures for conducting the 3D visual SLAM process. Optical flow is an image-processing theory that was proposed by Lucas and Kanade in 1981 and that is explained by Sonka et al.[12] The camera on a mobile robot is a moving camera that provides information such as motion parameters, consecutive image frames, relative distances of objects in the image, and a corresponding optical flow image. However, optical flow analysis does not produce motion trajectories, and only the important motion characteristics are detected. These general characteristics include mutual velocity, depth information, collision prediction, and interesting (i.e., feature) points, which can improve the reliability of complex dynamic image analysis and processing significantly. Translational motion means a mobile robot is moving relative to a SAR object. The constant light source carried by the robot supplies stable light for illumination. The depth translation movement of a mobile robot relative to the static objects forms a series of consecutive image frames. The optical flow changes corresponding to the translational motion, as shown in Fig. 1(a)–(c), where Fig. 1(a) shows the experimental setup used to obtain RGB image frames, this being installed on a robot with an industrial PC (IPC).

The mutual vector velocity between the robot and the OOI can be expressed by optical flow as follows:

$$(x', y') = \left(\frac{x_0 + c_x t}{z_0 + c_z t}, \frac{y_0 + c_y t}{z_0 + c_z t} \right) \tag{1}$$

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

Algorithm 1 Optical flow computation of object image during robot translation motion.

Input: image sequence frames in the translation motion process

Output: mobile robot position co-ordinate containing depth information

1. $c(i, j) \leftarrow c(i, j) = 0$; \\ initialize velocity vectors.
2. Points $(c_x, c_y) \leftarrow c(i, j)$; \\ evaluate starting values (c_x, c_y) of the optical flow.
3. Compute image velocity

$$c_x = dx/dt,\ c_y = dy/dt$$

4. Image co-ordinate calculation \\ x_0, y_0, and z_0 are initial coordinates in Eq. (1).

$$x' = (x_0 + c_x \cdot t)/(z_0 + c_z \cdot t),\ y' = (x_0 + c_y \cdot t)/(z_0 + c_z \cdot t)$$

5. Obtain mobile-robot position containing depth information:

$$x(t) = \frac{x'(t) \cdot c_x(t) \cdot D(t)}{dD(t)/dt},\ x(t) = \frac{y'(t) \cdot c_z(t) \cdot D(t)}{dD(t)/dt},\ z(t) = \frac{c_z(t) \cdot D(t)}{dD(t)/dt}$$

6. Set $d_{\min}(i, j) = \left\| \vec{x} - \vec{x}_{ob} \right\|_{dt \to 0}$, $r \leftarrow r_0$ \\ $d_{\min}$ is the smallest distance in Eq. (8)

for $(i, j) \leftarrow (1, 1)$ until n iterative times do
if radius $r > d_{\min}(i, j)$;
then repeat step 3;
else if stop;
end if;
end for;
return step 1;

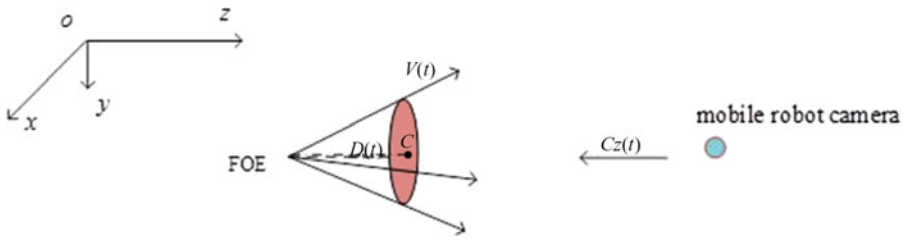

Fig. 2. Relationship between optical flow parameters of an image and the z-direction speed of a mobile robot camera.

where (x_0, y_0, z_0) are the coordinates of the OOI at initial time $t = t_0$, (c_x, c_y, c_z) are the mutual speeds in the x, y, and z directions, respectively, between the OOI and the mobile robot, and c_z contains depth information for 3D SLAM reconstruction. The image position (x', y') at time t is expressed by Eq. (1). The focus of expansion (FOE) of the 2D image can be determined by the optical-flow mutual velocity as

$$x'_{\text{FOE}} = \left(\frac{c_x}{c_z}, \frac{c_y}{c_z}\right) \tag{2}$$

In Eq. (1), the z coordinate can be obtained as $z(t) = z_0 + c_z t$. The presence of a z coordinate means that optical flow can be used to determine the depth information between an OOI and the robot. In Fig. 2, the red object denotes a ball image of Fig. 1, and $V(t)$ represents the image vector velocity in the optical flow. Let $D(t)$ be the distance from the FOE of the image's optical flow to a given point in an image frame measured using a 2D image; the given point is chosen as the intersection spot C, which is the endpoint of the $D(t)$ dotted line and is marked in both Figs. 1 and 2. This forms a basis formula for determining the depth distances between static OOIs and a moving robot, defined as

$$\frac{D(t)}{dD(t)/dt} = \frac{z(t)}{c_z(t)} \tag{3}$$

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

where $z(t)$ is the coordinate value of the object image in the z direction and represents the depth distance between the OOI and the mobile robot. Let the image vector velocity of the optical flow be $V(t) = dD(t)/dt$; the depth information of the OOIs can then be obtained from the optical-flow parameters of the image frames, as shown in Fig. 2.

Using Eqs. (1) and (3), the relationship between real-world coordinates (x, y, z) and image coordinates (x', y') can be found corresponding to the robot position and velocity as follows:

$$x(t) = \frac{x'(t)c_z(t)D(t)}{dD(t)/dt}, y(t) = \frac{y'(t)c_z(t)D(t)}{dD(t)/dt}, z(t) = \frac{c_z(t)D(t)}{dD(t)/dt} \tag{4}$$

The optical flow approach can be used to detect potential collisions with OOIs within the exploration area. As seen from the optical flow view, the mobile robot moves toward the FOE of this translational motion, and how the focal point of consecutive image frames proceeds can be expressed as

$$\vec{s} = \vec{i} \cdot \frac{c_x}{c_z} + \vec{j} \cdot \frac{c_y}{c_z} + \vec{k} \cdot 1 \tag{5}$$

The point coordinates of the mobile robot at each time t form a trajectory vector which is expressed as

$$\vec{x} = (x, y, z) = t \cdot \vec{s} = \left(\frac{t \cdot c_x}{c_z}, \frac{t \cdot c_y}{c_z}, t \right) \tag{6}$$

where c_x, c_y, and c_z represent the point velocity of image in the x, y, and z directions, respectively.

The location of an OOI in relation to the mobile robot's position $\vec{x}$ at time $t + dt$ is defined as

$$\vec{x}_{ob} = \frac{\vec{s} \cdot (\vec{s} \cdot \vec{x})}{\vec{s} \cdot \vec{s}} = \left(\frac{c_x}{c_z}, \frac{c_y}{c_z}, 1 \right) \cdot (t + dt) \tag{7}$$

The smallest distance $d_{\min}$ between the OOI's position $\vec{x}_{ob}$ and the mobile robot's position $\vec{x}$, during time interval Δt can be expressed by

$$\begin{aligned} d_{\min} &= \left\| \vec{x} - \vec{x}_{ob} \right\| \\ &= \sqrt{(\vec{x} \cdot \vec{x}) - \frac{(\vec{s} \cdot \vec{x})^2}{\vec{s} \cdot \vec{s}}} \\ &= \sqrt{2 \left[\left(\frac{c_x}{c_z} \right)^2 + \left(\frac{c_y}{c_z} \right)^2 + 1 \right] \cdot t \cdot \Delta t} \end{aligned} \tag{8}$$

where $d_{\min}$ represents the first equalities of the smallest distance between Cartesian coordinates of the robot position $\vec{x}$ and OOI position $\vec{x}_{ob}$; the second equalities calculates the smallest distance by using the Cartesian coordinates of image frames' focal $\vec{s}$ in Eq. (5) and the robot position $\vec{x}$ in Eq. (6); the last equalities obtain the smallest distance by using optical flow mutual velocity c_x, c_y, and c_z, the time t and Δt, which are defined in Eqs. (1), (2), and (7).

Therefore, to avoid collisions during exploration, a robot of radius r should satisfy $d_{\min} < r$.

Using Eqs. (1)–(8), the optical-flow processing steps are (i) image acquisition, (ii) object outline detection, (iii) extract the features (e.g., color, contours, texture, and shape) of searched OOI, (iv) optical flow computation, (v) object identification (OI), (vi) robot pose estimation, and (vii) robot localization itself while moving to the object. The OOI in step (iii) could be a trapped person, a victim, a dangerous objects, or an exit, among many others.

2.2. *RGB-D camera obtains depth information from image*

In USAR SLAM, the competing objectives (i.e., robot localization, mapping, exploration, feature distribution discovery, and searching for victims) are conducted using an RGB-D camera.[13] We developed a 3D reconstruction approach based on vision and graphics, which reduced a large quantity of point cloud data to a smaller quantity without compromising the visual quality of OI. Using a standard RGB-D benchmark data set, the real 3D environment thoroughly reconstructed from a 2D

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

Fig. 3. Visual system used in 3D reconstruction mapping. The system is shown performing digital processing on an image captured by an RGB camera.

map with a feature-based mapping method.[14,15] An RGB-D camera is a full 3D mapping system fusing visual features, optical flow signals, depth information, and shape-based alignment.[8] As shown in Fig. 3, the RGB-D vision used in our study is RealTime ICETEK, a digital signal processing system adapted for robots performing SLAM autonomously.

The RGB-D camera system captures image depth information, while 3D data are represented by stereo coordinates in space; the depth information is the z coordinates of the distance from the spatial object to the camera. RGB-D camera mapping provides depth information only within a 5-m radius. LRF scans are usually employed for 3D mapping, and their view field is even wider (e.g., 0°–180°).[8] RGB-D mapping results in a full 3D structure combining visual features and shape-based alignment. Visual and depth information are combined to complete loop-closure detection. The lens of RGB-D camera is an analog low-pass filter in the spatial domain, In other words, the illumination frequency of the environment cannot be too high because the camera can capture only low-frequency band light.[16] The cutoff frequency u_0 is generally represented as

$$u_0 = D/\lambda f_0 \tag{9}$$

where D is the diameter of the lens aperture, λ is the optical wavelength, and f_0 is the focal length of the camera. The aperture is used to control the light passing through the lens into the body of the light-sensitive surface, which is usually in the lens internal.

Our method employs a single RGB-D camera sensor mounted on a robot to extract image-feature depth information for 3D map reconstruction. The mathematical model for depth is developed based on the HSV (hue, saturation, value) color space map,[17] and the quadratic calibration equation is

$$d = 0.0032\,H^2 - 1.9492\,H + 359.26 \tag{10}$$

where d is the converted estimated distance based on the image's hue value H. The coefficients in Eq. (10) were determined from simulated experiments. Figure 3 shows the experimental setup used for the digital processing of the RGB-D camera images, and Table I lists the depth distances estimated according to the hue values of image RGB color. The coefficients were determined from Eqs. (11), (12) and experimental data given in Table I.

Equation (10) is an experiment-based mathematical model for computing depth distance from the hue H, while the HSV theory of RGB is given by Eqs. (11) and (12), H' is the theoretical coefficient and the RGB hue values are transposed according to $H = 60^0 \times H'$, that is, creating the mathematical model of Eq. (10) based on measurements and theoretical values H. A robot-mounted Kinect sensor captures RGB-D images for 3D-reconstruction mapping while exploring a disaster area. A stereo RGB camera in which a single RGB-D camera generates environmental 3D-reconstructions within its visual range and extracts image depth information for conducting SAR SLAM. This method reduces the cost and calculation time for SLAM.

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

Table I. Depth estimated from Hue value of image RGB color.

Sample	R(°)	G(°)	B(°)	Hue value H(°)	Estimation depth d (cm)
1	20	36	48	200°	97.42
2	0	60	255	225.88°	82.24
3	2	162	255	202.06°	96.06
4	0	249	255	181.41°	110.97
5	0	255	31	127.29°	163.00
6	20	48	36	160°	129.31
7	48	36	20	360°	72.27
8	36	20	48	280°	64.36
9	36	48	20	80°	223.80
10	48	20	36	320°	63.20

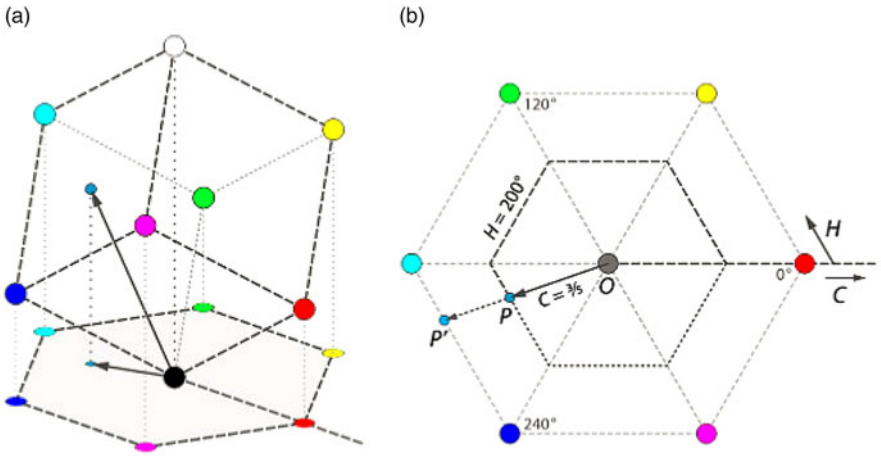

Fig. 4. Hue values of RGB colors represent the image depth. (a) Hue and chroma. (b) A sample point P': $R = 1/5$, $G = 3/5$, $B = 4/5$, $C = M - m = B - R = 0.6$, $H = 60^0 \times [4 + (R - G)/C] = 200^0$.

The common-sense depth concept in V-SLAM is the distance from the mobile robot to the searched OOI. Because the R, G, and B components of an object's color in a digital image are all correlated with the amount of light hitting the object, the HSV version of RGB is used in robot vision for image analysis and OI. Combining HSV version of RGB17 and the Joblove and Greenberg (J–G) equation, whereby HSV values were first introduced, the depth information is derived from the image by converting the RGB color to hue values that can be used to estimate the relative distances between a robot and the surrounding targets in a SAR environment. According to the J–G equation,[17] the transposed hue value is expressed as

$$H' = \begin{cases} 0°, & M = m = R = G = B,\ C = 0; \\ mod.\left(\frac{G-B}{C}\right) \le 6, & M = R,\ C = R - m; \\ \frac{B-R}{C} + 2, & M = G,\ C = G - m; \\ \frac{R-G}{C} + 4, & M = B,\ C = B - m \end{cases} \tag{11}$$

where $M = \max(R, G, B)$, $m = \min(R, G, B)$, and $C = (M - m)$, and

$$H = 60^0 \times H' \tag{12}$$

where H is the hue value of the RGB image pixel. When $R = G = B$, that is, $C = M - m = 0$, it is neutral and the transposed hue value is assigned as 0°.

The average values of the object's RGB image are extracted and converted to a hue value according to Fig. 4(a), (b). The relationship between the distance and hue value is given in Table I, and the data

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

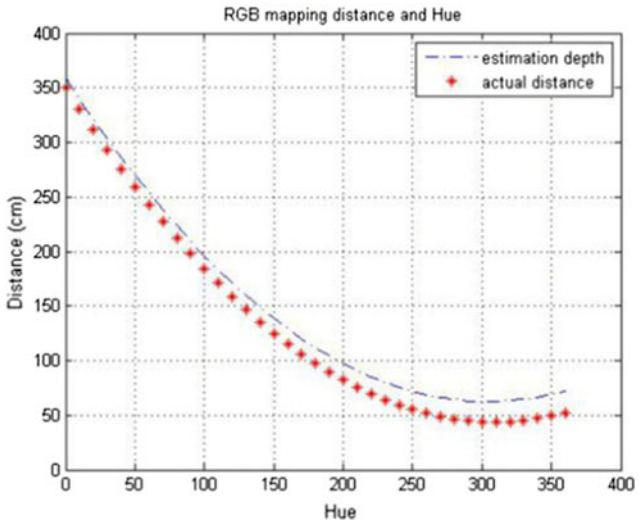

Fig. 5. Comparison of object's actual distance with that estimated from image RGB hue values.

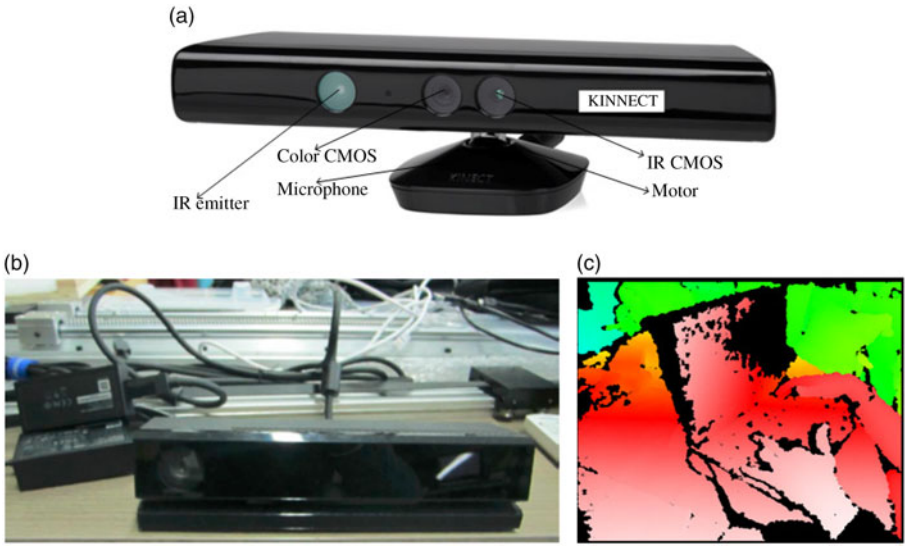

Fig. 6. Kinect visual system for capturing RGB images. (a) Components of Kinect sensor. (b) Kinect-style monocular RGB-D camera. (c) Depth map is analyzed visually using colors from white (near) to blue (far).

in last column represent estimated depth measurements. Table I is the digitalized data of Eq. (10) and J–G equations (11) and (12), and the pixels in consecutive images are shown in Fig. 6(c).

The hue value of the RGB image provides an estimate of the depth according to Eqs. (10)–(12). In Fig. 5, the depth estimated from the hue value of the RGB image is compared with the actual distance from the moving camera to the static OOI. These results indicate that using the hue values of the RGB image sequence is an effective way to estimate the depth information in 3D scenario reconstruction.

2.3. Kinect visual system for capturing RGB images to obtain depth information

An intuitive user interface is used for mobile robot exploration in a real or virtual rubbly disaster environment. However, it is extremely difficult to extract dense depth information from camera data alone in dark or sparsely textured environments.[8] Microsoft released the Kinect Software Development Kit for Windows 7 on June 16, 2011, and we developed a Kinect 3D scanner system to employ RGB camera to capture the scene's depth information.

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

Table II. Image sharpening function program.

Input	void sharpen2D (const Mat &image, Mat &result)
Output	sharpened image, result, image.depth(), kernel
Start	
Mat	kernel ← (3, 3, CV_32F, Scalar(0));
	(i, j) ← (0.0); // construct a Matrix kernel, and initialize each item
	kernel.at (0,1) ← (–1.0);
	kernel.at (1,0) ← (–1.0);
	kernel.at (1,1) ← (5.0);
	kernel.at (1,2) ← (–1.0);
	kernel.at (2,1) ← (–1.0); // assign value to each kernel item of Matrix
	filtered2D (image, result, image.depth(), kernel); // sharpening filtering the image
end	

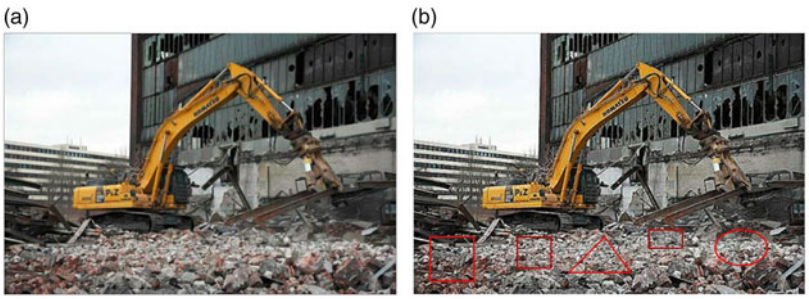

Fig. 7. A building demolition scene simulating a SAR scenario for a mobile robot collecting SLAM data. (a) Original scene. (b) Image after sharpening processing to highlight interesting objects (i.e., the areas enclosed in red shapes).

As shown in Fig. 6(a), (b), the depth sensor consists of an infrared ray (IR) projector, a monocular RGB camera, and an IR complementary metal-oxide-semiconductor (CMOS). Figure 6(a) shows the Kinect system that is used to process the relationship between RGB images and the depth information. In Fig. 6(b), the Kinect-style monocular RGB camera includes an infrared laser projector combined with a monochrome IR CMOS sensor so that it can capture 3D video image sequences under any ambient light conditions. Figure 6(c) shows the depth map visualized using the RGB-D representation, that is, from white (near) to blue (far). The sensing scope of the depth sensor is adjustable, and the Kinect software can calibrate itself according to the surrounding objects or other obstacles. Kinect's various sensors output video at a frame rate of 9–30 Hz according to the resolution. The RGB sensor achieves 8-bit video graphics array (VGA) resolution (640 × 480 pixels), and the monochrome depth-sensing video stream achieves 16-bit resolution (640 × 480 pixels).

Using the open-source computer-vision library OpenCV to manipulate images and features, a sharpening spatial filter reconstructs environmental scenarios to highlight characteristic features of OOIs. Figure 7(a), (b) shows a cluttered demolishment scene that was processed by the proposed RGB-D algorithms to highlight OOIs in SAR environments, and a sharpening-filtered image was obtained. Figure 7(a) shows the initial scene viewed by a mobile robot collecting SLAM data, and Fig. 7(b) shows the resulting image processed by the sharpening program proposed in Table II. The image sharpening method is developed to reconstruct the map that can highlight the OOIs, and the sharpening filter matrix is established according to

$$M_{sh} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{13}$$

where the center element of the matrix corresponds to the current pixel of interest. Different sharpening filters generate different reconstructed images. The sharpening process depends on a kernel

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

matrix, which defines an image filter. The kernel matrix is shown in Eq. (13); it contains a 3×3 neighborhood of pixels. The output of such a neighborhood sharpening filter includes the following compositions: (i) the two horizontal pixels are each multiplied by -1; (ii) the two vertical pixels are each multiplied by -1; (iii) the pixel of interest is multiplied by 5.

OpenCV is employed as robots' vision system to detect and identify RGB objects by running a C++ image processing program (http://opencv.org/). In the developed 2D/3D objects reconstruction SLAM, a mobile robot processes RGB-D images through the OpenCV approach. Equation (13) is a typical filter function of OpenCV, and this mathematical model highlights the areas, objects, and key features of interest by choosing different optical filter algorithms. Algorithm 2 illustrates the meaning of each filter element value; for example, the maximum value represents the area of greatest interest.

Algorithm 2 Image processing of neighborhood sharpening filter.

0	Pixel value $\times(-1)$	0
Pixel value $\times(-1)$	Pixel value $\times$ 5	Pixel value $\times(-1)$
0	Pixel value $\times(-1)$	0

The optimal algorithms can segment sub-maps to improve the identification of interest. Table II represents the code program of OpenCV. The simulated SAR scene in Fig. 7 is processed using OpenCV procedure to validate the algorithm, and the processed image contains highlighted areas enclosed in red triangles, rectangles, and circles. This sharpening filtering process is shown in Algorithm 2. For a particular kernel matrix, the image-sharpening program is represented by Table II.

Equation (13) is the proposed main representative approach with OpenCV to process RGB-D image, and the other more complex methods can do so according to these steps. The methods are elaborated in Algorithm 2 and Table II. RGB-D image depth information is developed for the proposed 3D reconstruction of SAR objects on the basis of the 2D SLAM process. Algorithm 2 contains the developed image processing filters for OI in 3D reconstruction, and Table II displays one of procedures for conducting the proposed SLAM process.

2.4. Topological graph reconstruction based on image fusion

The proposed 3D visual SLAM is an OI method to build sequential 2D/3D maps within a visual SLAM framework in cluttered USAR environments. An efficient 3D scene reconstruction plays significant roles in robotic SAR post-disaster scenarios; a vision-based sensory package and an extended information filter (EIF)-based SLAM algorithm have been designed.[14,18] Probabilistic methods, 3D scans, and the ICP algorithm are existing techniques for SAR scenarios.[19]

Topological graph-based methods are developed for data matching and detecting OOIs and loop closure (LC). In our method, mobile robots conduct 6-DoF SLAM, which considers six dimensions for the robot's position: three Cartesian coordinates and the angles of orientation (i.e., roll, yaw, and pitch). The method is based on ICP scan matching, and the initial pose estimation employs a coarse-to-fine strategy with an octree representation and loop-closing detection. A scale-invariant feature transform (SIFT)-based approach incorporates both 3D depth images and 2D imagery from a structured light sensor. Digital fringe projection and a phase-shifting technique provide online 2D and 3D sensory information at a frame rate of up to 60 frames per second (fps). Landmarks are perceived in the images by using a combination of SIFT key points, depth segmentation, edge detection, and morphological techniques according to a convex hull algorithm.[20–22] An incremental approach to creating topological segmentation for unstructured environments was described in ref. [23]. This approach is based on spectral clustering of an incremental generalized Voronoi decomposition of discretized metric maps. It extracts sparse spatial depth information from the images and constructs an environmental model. The formed topological map is globally consistent in robotic USAR.

Multiscale transforms are popular image fusion methods, and the fused image is obtained by taking an inverse multiscale transform.[24] We employ a robot equipped with an RGB-D camera to

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

integrate 3D information and color video images into the map building, which produces a 3D virtual reconstruction of the USAR environment as the robot explores an area.[9] A 3D sequential map of visual natural landmarks and the mobile robot's location are obtained using a top-down Bayesian method to model the dynamic exploration trajectory.[25]

As Bloch described in ref. [26], the mobile robot's perception is achieved by the sensors embedded in the robot itself. The LRF installed in the mobile robot is used to obtain the range data of the rubble around the robot. The proposed processing of the information obtained from the sensors is arranged as follows: (i) data exploration according to a sequence of image frames using a dynamic model; (ii) focusing on certain elements that are more important than others; (iii) sensor-based control that is defined on the basis of the sensor's measurements. The 3D SLAM process is achieved by fusing data obtained from the sensors, the SLAM simulations and experiments, and a laser-type device mounted on the mobile robot that provides information regarding the distance between the robot and the OOIs.[27,28]

3. Three-Dimensional Visual SLAM Reconstruction Algorithm Adapted to SAR Scenarios

3.1. Wavelet transformation relates 2D and 3D images

The traditional Fourier transformation converts the image signal from the time domain to the frequency domain. Instead, we proposed wavelet transformation which is an improvement on the Fourier transformation; it transforms an image signal into both the time and frequency domains, which is necessary for 3D image reconstruction. A state-space model combining Rao–Blackwellized (R–B) particle filter performs the data association of assigning time-of-flight measurements to corresponding landmarks. A bat-type ultra-wideband (UWB) laser radar is applied to reconstruct a SAR environment, and different target objects can be used to distinguish and locate those features for SAR SLAM exploration.

Using N particles requires approximately N times more computational power than the simpler nearest-neighbor search (NNS) algorithm; however, the number of particles can be reduced in the R–B method, so it is still fast enough for real-time reconstruction of the environment.[29] In ref. [30], a mobile robot was used to explore and classify terrain in unstructured and natural environments based on range data provided by a stereo RGB-D camera system. A convolutional neural network was designed to learn autonomously which features are relevant to the searched objects.

Given a function $\psi(t) = L^1(R) \bigcap L^2(R)$, and $\hat{\psi}(0) = 0$, that is, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$, then $\psi(t)$ is a basic wavelet, and the primary principle of wavelet transformation is defined by Eqs. (14)–(18):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), a, b \in (-\infty, +\infty), a \neq 0 \tag{14}$$

The corresponding wavelet transformation formula is expressed as

$$WT_f(a, b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt = \langle f, \psi_{a,b}\rangle \tag{15}$$

and

$$WT_f(a, b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt = |a|^{-1/2} f^*(t)\bar{\psi}_{|a|}(b) \tag{16}$$

If

$$c_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\psi)|^2}{|\psi|}d\psi < +\infty \tag{17}$$

then $\psi(t)$ is referred to as an admissible wavelet, and the inverse transformation is expressed as

$$f(t) = \frac{1}{c_\psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{a^2} WT_f(a, b)\psi_{a,b}(t)dadb \tag{18}$$

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

The discrete wavelet transformation is applied for digital image processing and the parameters are set at $a = 2^{-j}$, $b = k2^{-j}$, where j is the scale parameter and k is the shift parameter, both of which are integers. From $a, b \in (-\infty, +\infty)$, $a \neq 0$ in Eq. (14), it can be shown that $i \in Z$ and $k \in Z$.

3.2. Improved ICPs for data association

Map building requires estimation of the robot's trajectory in 3D space. However, it is difficult to estimate the trajectory using odometry or gyro in an environment with collapsed rubble. The ICP algorithm is a fast and robust method for 3D mapping online. A compute unified device architecture (CUDA)-enhanced fast NNS approach is applied for ICP with a singular value decomposition (SVD) solver algorithm, which maintains the data registration's robustness. A convergence radius exploration ICP algorithm is an automatic system for gauging and digitalizing 3D environments.[31] The 6-DOF SLAM includes the following six dimensions: the mobile robot's 3D position coordinates (x, y, z) and the angular orientation data (yaw, pitch, roll), and an extended Kalman filter (EKF) is used for sensor fusion, neural field dynamics used for behavior generation, and evolutionary algorithms used to optimize each part of the system.[32–35]

We propose the novel IICP algorithm that uses visual features and their associated depth values to obtain an initial alignment. A SIFT detects the features, random sampling consensus (RANSAC) matches the data, and least-square is used to fit 3D points. The exploration information from an inertial measurement unit (IMU) improves the performance of the data association algorithm. An EIF is used to estimate the state vector containing the sequence of mobile camera poses and some selected 3D-point-cloud features in the environment. A range imager and a conventional camera obtain data information for building dense 3D maps in SAR environments. In our designed system, 3D LRFs mounted on mobile robots enable the automatic acquisition, surveying, and mapping of the entire SAR environment. This system combines the precision of the ICP-based methods and the flexibility of localization. Using this theory, neural networks, and the particle filter ideas, we developed 2D/3D SAR SLAM for robots autonomous exploration.

IICP registration and visual saliency-based presence detection enable a large portion of the image frames without a salient object to be ignored, which is optimized by the NNS algorithm; this procedure is designed as shown in Algorithm 3. Here, P_s is the RGB-D point cloud source, F_s computes the feature-point distance in P_s, P_t is the RGB-D object image frame, F_t computes the feature points distance in P_t, T_p is the original transformation, T^* is the corresponding optimal transformation, A_f indicates the searched layers of iteration, r is pre-setting minimal numbers, O represents empty and end, A_d means feature point associations, and δ is a small threshold of transformation.

Algorithm 3 Improved ICPs optimized by NNS method.

```
Start
F_s ← Extract_RGB_Point_Features (P_s);
F_t ← Extract_RGB_Point_Features (P_t);
(T*, A_f) ← Perform_RANSAC_Alignment (F_s, F_t);
if |A_f| < r then
   T* = T_p;
   A_f = O;
end
if |A_f| ≥ r then
return T*;
else
   repeat
   A_d ← Compute_Closest_Points T*, P_s, P_t;
   T* ← Optimize_Alignment T*, P_s, P_t;
   until (change T* ≤ δ) or (Iterations > Max Iterations);
   return T*;
end
```

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

3.3. Identifying OOIs with improved EKF during the SLAM process

EKF is a general model of probabilistic particle filtering, and we propose this mathematical method improved by joint probabilistic data association (JPDA) in 3D OI for 2D mapping. When more advanced data association methods such as probabilistic data association (PDA) or JPDA are used, the SLAM performance is improved regarding object detection in spite of the densely cluttered scenario. To solve the SLAM problem whereby its computational complexity increases with the number of landmarks, we developed improved dimensional-bounded EKF that can achieve good estimations with constant complexity when removing landmarks from the current state. The current challenge is to store old features, instead of removing them from the current state.[36]

In general, data association is considered to be a difficult task in robotic SLAM, especially when performing SAR SLAM in unstructured disaster environments. JPDA considers any measurement that originated from the cluttered scenario and that lies in the validation scale of the objects. PDA was developed to resolve the ambiguous data association by using a machine learning algorithm.[37] This algorithm is based on a linear discriminant classifier to enhance the data association strategy for performing SLAM tasks in a collapsed scenario, and measurements and targets are divided into different clusters. An improved feature-based 3D SLAM approach was described in ref. [38]; it involved a mobile robot equipped with a rotating laser scanner using the improved EKF to build a 3D map of the environment. The scanner sensor delivers dense 3D point clouds and tracks the mobile robot's pose.

3.4. Transformation from 2D images to 3D point clouds for depth map reconstruction

Following Santos et al.[6] and Carbone et al.,[39] we employ a task-based attentive exploration system designed for autonomous robots working in SAR scenarios, in which the main tasks are to explore the area, find victims, and identify targets using an autonomously built map. The specificity of the rescue scenario is usually based on shape and color features, and the developed 3D visual-attention process guides the robot's exploration through an incrementally generated viewpoint saliency map. The exploration strategy is the result of the cooperation between perceptual attention and SLAM. Various pre-attentive methods, such as the robot determining the weight and significance of salient sources in the scene, are used to consider pre-attentive feature selection useful for attentive exploration.

The saliency values represent weight factors of OOIs in the global metric map and are extracted from the perceptual features by analyzing the visual scene; these values are integrated into a global saliency 2D map superimposed over the metrical one. This map identifies interest viewpoints that can be used to guide a task-oriented exploration. However, a pure 2D metric exploration involves the mobile robot exploring a SAR area by using the initial standard metric map while simultaneously, updating the original metric map by producing the new map of the exploration environment. Different from this metric exploration, the proposed task-based attentive exploration significantly improves exploration effectiveness by using a saliency-driven method, where saliency-driven means that the mobile robot is driven by the saliency values of the global metric map. Additionally, extending the map building by including multimodal integration of different perceptual features instead of only visual ones yields substantial improvement.

Features extraction and the creation of point clouds are the proposed novel 2D SLAM, while depth information is responsible for the reconstruction of 3D OOIs. According to Marr's theory,[12] 3D visual reconstruction can be built from attentive object images, which results in a 3D model of a SAR scenario with very little *a priori* knowledge about the objects. Nonhomogeneous objects represented by 2D points $[u, v]^T$ can be mapped to a nonhomogeneous image point $[u, v]^T$ by the transformation matrix H, which is shown as

$$\alpha \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{19}$$

where $u' = \frac{h_{11}u+h_{12}v+h_{13}}{\alpha}$, $v' = \frac{h_{21}u+h_{22}v+h_{23}}{\alpha}$, and $\alpha = h_{31}u + h_{32}v + h_{33}$.

Adjustments to the scale parameter α can yield better images of objects in a robot exploration area. Schleicher et al.[25,40] managed 3D characteristics using a multilevel relaxation (MLR) algorithm that improves the pose estimation from the RGB-D visual sensor system by evaluating the addition of an IMU.

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

Because matching a highly topological 2D map with local-level 3D objects' maps causes degradation of the map-building accuracy, a real-time SLAM method based on only stereo vision is employed for large-scale outdoor environments, and the MLR algorithm manages only 2D information. The robot's 2D position X_{2D}^{fp} is defined as

$$X_{2D}^{fp} = (x_{2D} \quad y_{2D} \quad \theta_{2D}) \tag{20}$$

and the relationship between the 2D and the 3D pose X_{3D}^{fp} is expressed as

$$X_{3D}^{fp} = \left[x_{2D} \quad z_{2D} \quad 2\arccos\left(\alpha_{0_{x_{2D}}}^{y_{2D}}\right)\right]^T \tag{21}$$

where x_{2D} represents x directional coordinates of the 2D map, y_{2D} represents y directional coordinates of the 2D map, and θ_{2D} indicates the angle between the robot's motion trajectory and x axis, z_{2D} represents z directional coordinates of reconstructed 3D object, and $\alpha_{0_{x_{2D}}}^{y_{2D}}$ is the angle between reconstruction 3D map of object between global 2D map planar.

Sparse visual features might not be captured in dark areas. To overcome this limitation, RGB-D IICP combines RANSAC and dense cloud point alignment. A sequential window-based bundle that considers multiple frames at a time improves the alignment accuracy.

3.5. Information fusion in map reconstruction

Information fusion is a common research means. As described in ref. [35], data association is performed using a combination of SIFT feature detection, RANSAC matching, and least squares fitting of 3D point sets. For a USAR robot exploring a collapsed scenario, the highest priorities are marking exit points, stairways, windows, and large cavities on the map of the building. The EIF algorithm not only produces consistent SLAM outputs but also incorporates dynamically into the estimation process for efficient 3D exploration of collapsed terrain. There is a trade-off between computational efficiency and information loss when fusing the minimal information generated based on the robot's uncertain belief and the perceived quality of the observation.

Following ref. [41], we employ a conventional camera and a range imager in 3D point-based SLAM to provide range, bearing, and elevation inputs to visual salient features, which contribute to 3D texture-rich maps. A visually selective and attentive method can efficiently recognize objects in an exploration area, and visual information from a robot's onboard camera can provide rich sensor data.[42] A line-matching localization and mapping algorithm extracts the lines from the LRF based on a least squares algorithm.[43] The benefit of using lines instead of points is that it reduces the computational intensity dramatically. ICP scan-matching EIF algorithms make it possible for a robot to perform accurate positioning in unstructured environments. As described in ref. [44], we designed a perception-driven exploration algorithm for visual SLAM, which balances exploration and revisitation automatically. A multilayer, hybrid, distributed field robot architecture, which successfully integrated operator graphical interfaces with high-level deliberative and low-level reactive robot services, and a visual display provide depth information of OI.[45,46]

In our method, the display shows the vital information based on the following procedure: (i) extract the distance data between the object and the robot as represented by each pixel when the object is within a predefined scope; (ii) track the object in the visual display; and (iii) display continuously the distance from the robot to the selected object.

4. SLAM Simulation on the MobileRobot Platform Using Virtual Robots

We developed novel robotic SAR SLAM methods using the MobileRobot platform, and we tested the 2D/3D reconstruction V-SLAM using virtual robots, Amigo mobile robots, a Pioneer LX robot, and a crawler LUKER robot in outdoor and indoor simulated SAR scenarios. MobileRobot is an open source platform of robotic control, and its environment comprises a modified simulation in which virtual p3dx mobile robots are controlled via TCP port 8101.

As shown in Fig. 8(a), (b), the virtual mobile robots explore and map the loaded metric map. Virtual mobile robots conduct SAR SLAM program on a MobileRobot platform with a demonstration robotic arnlServer. MobileRobot uses the MobileEyes, MobileSim, and Mapper3 software to monitor and control the robot and its accessories (e.g., RGB-D cameras, LRF, sonar, and gyroscope). With this

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

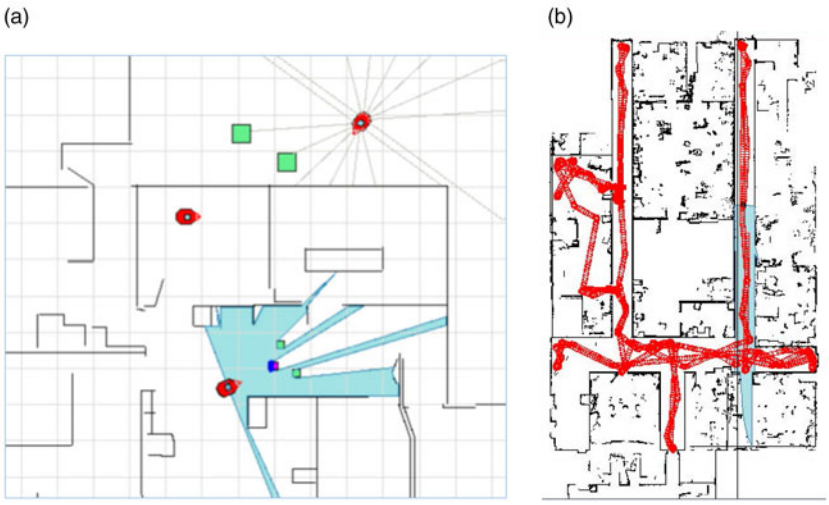

Fig. 8. Virtual robots mapping the simulation environment. (a) Gray lines and blue bands indicate sonar and laser scanning. (b) The exploration trajectory (red band).

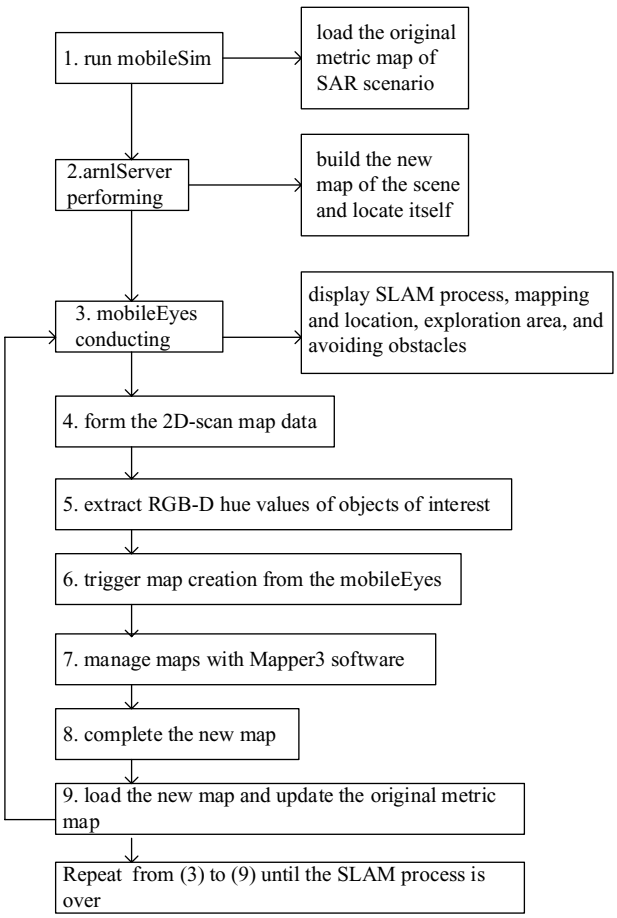

Fig. 9. Flowchart of SLAM simulation with MobileRobot-simulated platform.

simulated platform, the virtual mobile robots can run all programs to conduct localization, control, mapping, SLAM, and exploration from the loaded metric map of SAR environments. The SLAM simulation process is performed according to the flowchart as shown in Fig. 9.

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

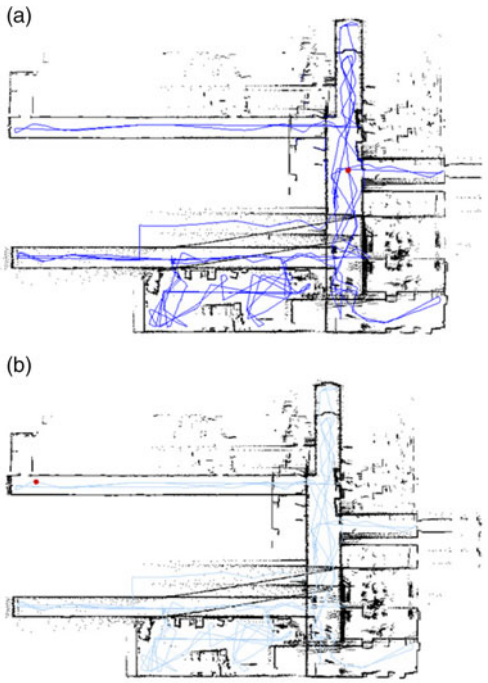

Fig. 10. Display of virtual mobile robot building a map and localizing itself simultaneously. (a) 2D scan data and RGB-D hue values (the red dot represent the mobile robot. (b) Processing the acquisition data to form a new map.

In Fig. 10(a), (b), the 2D scan data and RGB-D hue values for depth information are acquired in the process of the mobile robot's exploration. By processing these relative data, a new environmental map is formed. Using mapping simulation software, the SAR OOIs and keyframe features are located and marked simultaneously in the produced new map. Finally, the current original metric map is updated online by the new produced map; the mobile robot then repeats the SLAM process. The complete SLAM process is shown in Fig. 11(a), (b), where the OOIs of RGB-D areas, triangles, and rectangles are marked on the 2D map.

5. Experiments with Robot SLAM Adapted to SAR Environments

The proposed 3D SAR OI on 2D SLAM tasks with the developed algorithms is tested using Amigo mobile robots, a Pioneer LX onboard robot arm, and a crawler LUKER robot. The flowchart of the SLAM procedures is shown in Fig. 9 (Steps 1–9). The methods for RGB-D object reconstruction are tested on Amigo robots in Fig. 1, where the red ball is taken to be an OOI. Consecutive images are captured by running the proposed Algorithm 1 according to Eqs. (1)–(9). Figure 2 shows the depth computation of 3D reconstruction. With RGB-D camera system installed on mobile robots, Fig. 4 shows the process of conducting the 3D object reconstruction SLAM according to Eqs. (10)–(12). Table I shows a data set of 3D depth measurements based on RGB color analysis. In Fig. 5, a Kinect-style RGB-D visual algorithm is designed according to Algorithm 2 and is placed onboard a crawler LUKER robot as shown in Figs. 13(a), (b) and 14(c), (d). The used filtering algorithm is Eq. (13), and the corresponding OpenCV program is run as Table II. The image shown in Fig. 7 is used to test the proposed filtering algorithms. The error analysis of comparison is given in Figs. 5 and 14 and shows the generated 2D map marked with colored OOIs, and Fig. 14(f)–(h) shows the reconstructed 3D maps of the SAR environment.

A Pioneer LX onboard the robot vision system is used to test 2D/3D objects reconstruction SLAM, Fig. 12(a)–(f) shows the indoor cluttered experiment. The Pioneer LX mobile robot conducts the SAR SLAM process procedures and then obtains 2D scan data and RGB-D hue values for depth information. Consequently, an LC is completed as shown in Fig. 12(c); the new updated is formed with RGB image marked goal1, goal2, and goal3 as shown in Fig. 12(f). The IPC installed on the

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

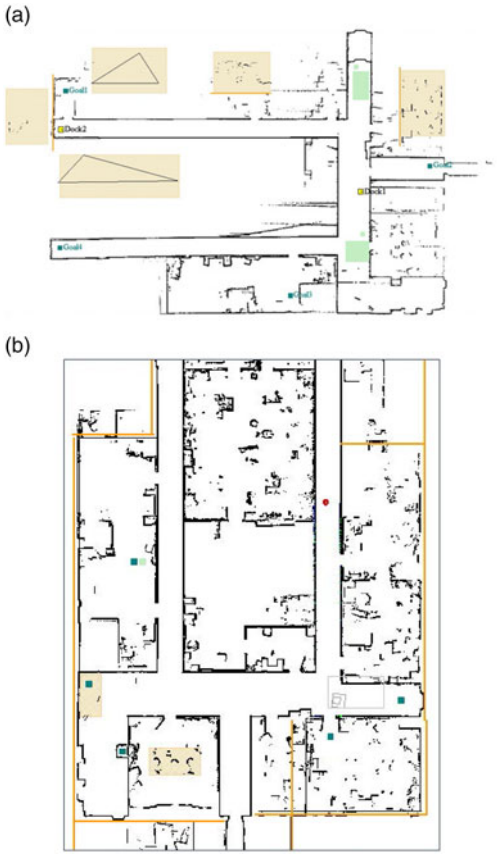

Fig. 11. New updated map produced from continuous SLAM process. (a) OOIs located and marked on new map. (b) Current original metric map updated online by the produced new map, which is corresponding to the actual current SAR environmental map.

Amigo mobile robot is an MXE-1300 Series, Intel Atom D2550/N2600 Fanless Embedded Computer with Integrated I/O (ADLINK Technology Inc.).

In Fig. 12(a)–(f), the mobile robots explore an artificial cluttered laboratory environment and a large-scale collapsed area; a complete updated map and the localization process are shown. Figure 12(b) shows the arnlServer SLAM program running on the IPC-equipped mobile robot that can be monitored by remote PC control center. This indicates that robots performing autonomous SLAM at a SAR scene can be controlled by remote PC control center, thereby ensuring the safety of robots.

LC generally means detecting and identifying the revisited place (a landmark or a salient feature), and Fig. 12(e) shows a perfect closed-loop detection by identifying a red round mark.

A crawler mobile robot is used as an experimental robot in unstructured SAR areas, and its components are shown in Fig. 13. The data used in the experiment were collected in a gypsum mine collapse disaster. A mobile robot is equipped with a rear RGB-D camera for observing the tether and cable; the cable guard is modified to prevent debris from falling into the device. A LUKER joint-type crawler mobile robot performs SAR tasks as shown in Fig. 14.

Figures 13 and 14 show outdoor unstructured SAR SLAM experiments with a LUKER crawler robot equipped with an RGB-D camera system, an LRF and gyro sensors. Using this robot, the SLAM experiments were conducted as shown in Fig. 14. The crawler robot performed SLAM tasks in a gypsum mine collapsed scenario; this was a gypsum-mine SAR scene as shown in Fig. 14(a), (c) and in a cluttered demolition environment as shown in Fig. 14(b). Figure 14(d) shows the robot performing SLAM process in an unstructured area, and the newly produced updated map is shown in Fig. 14(e) marked with OOIs 1, 2, and 3. Our 2D/3D object reconstruction methods combined the

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

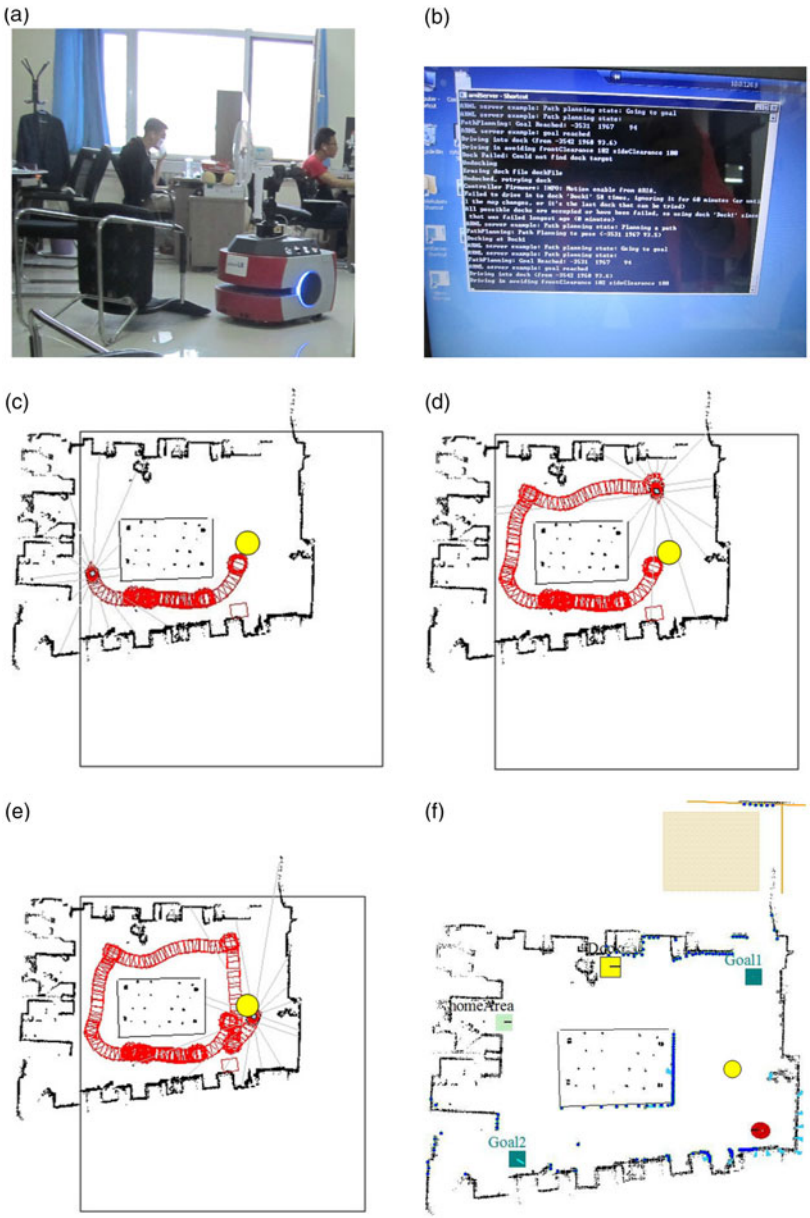

Fig. 12. Robot conducts SLAM process in an experimental indoor cluttered environment. (a) Experimental Pioneer LX mobile robot. (b) ArnlServer program is running. (c) Initial visited object. (d) Middle stage of exploration trajectory. (e) Revisited OI and complete loop-closure exploration trajectory. (f) Mobile robot explores the experimental cluttered area and a new updated map is generated.

advantages of 2D SLAM (i.e., its simplicity and speed) and the performance of 3D OI. Figure 14(e)–(h) shows 3D reconstruction of this exploration environment made using the installed RGB-D Kinect and the localized sensors.

6. Results and Discussion

The main contribution of this paper is the reconstruction of 3D SAR OOIs on the basis of a 2D SLAM process adapted to post-disaster scenarios. For this purpose, an autonomous robot system was built, comprising a wheel-driven platform, a gimbaled 2D/3D LRF, an IR camera installed in Kinect, and several RGB cameras. The mobile robot explored a cluttered experimental area using an original metric map that was preset on the robot. The robot avoided static and dynamic obstacles that could change while the robot was exploring. The robot started from different initial positions and avoided obstacles during the process of perception exploration. Based on the map that was produced

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

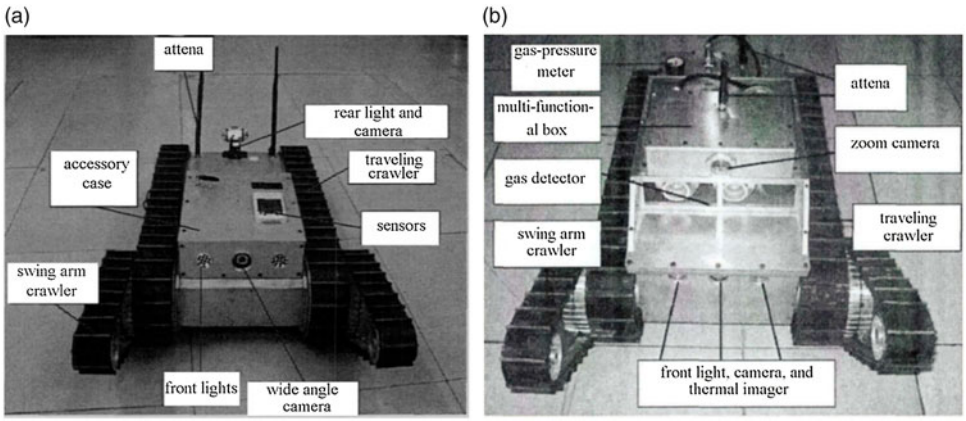

Fig. 13. Components of crawler search and rescue (SAR) mobile robot. (a) Top view. (b) Front view.

by fusing the 2D occupancy grid, the information about the dynamic obstacles was detected from 3D scans and the current LRF scan. In the exploration phase of the SAR process, the original metric map is constantly compared with corresponding updated map.

Comparing with the existing methods in the literature, the proposed 2D/3D reconstruction SLAM located and marked the SAR OOIs by planar segmentation local sub-maps as in the simulations and experiments in the corresponding Figs. 7(b), 11(a), (b), 12(f), and 14(e). As described in ref. [47], the map re-corrections are made when an obstacle is detected. Combining a localization of visual sensor and odometry provides a cost-effective solution for performing SAR SLAM in an unstructured post-disaster scenario.[36] The distance measurements provide localization in both the 2D and 3D cases.[48–50] The S-PTAM approach and LIBELAS algorithms used in ref. [51] are traditional SLAM method, which can be found in its cited references; moreover, the approach is absent in the robots experiment validation. Reference [52] reviewed Structure from Motion (SfM) and V-SLAM techniques. Comparing with the usual SLAM used in literatures such as refs. [51, 52], the 2D/3D SAR SLAM methods we proposed herein is adapted to uncertain environmental SLAM. With the proposed methods, Fig. 15(a) and (b) show the significant performances of 3D OI and LC detection on 2D SLAM process comparing with Fig. 14(e) and comparing with Fig. 12(e).

In simulations and experiments, RGB color marks [Fig. 7(b)], areas, red triangles, blue rectangles, yellow circles [Figs. 7(b), 11(a), (b), 12(f), and 14(c)], and graphics have validated the novel 2D/3D SAR SLAM methods. The 3D visual SLAM is a landmark identification method. A 3D real-time mapping sensor is used to build sequential 3D maps within a visual SLAM framework in cluttered USAR environments. The developed 3D visual-attention process guides the robot's exploration through an incrementally generated viewpoint saliency map. Loop closure generally means detecting and identifying the revisited place (a landmark or salient feature). An optimal LC strategy maintains independence of the 3D local map of objects; simultaneously, the 2D topological map maintains global consistency. A series of independent local 3D maps is associated with the global topological 2D map. SIFT descriptors match and store the graph-pose scheme, and key frames are used to decide whether the mobile robot is revisiting a place or not, that is, whether a loop is being closed.[40] Figure 3 compares mobile robot trajectories between a wandering exploration and an LC exploration in an uncertain environment. Once an LC has been detected, the whole map is corrected according to the recognized location. Figures 15(b) and 12(e) compare mobile robot trajectories between a wandering exploration and an LC exploration in an uncertain environment. Figure 12(e) has implemented a perfect LC identifying detection.

In general, a gray scale image cannot be used to provide depth information in 3D object reconstruction with the 2D SLAM process. According to Fig. 6(c), the SAR OOIs are sampled by the RGB-D camera and Kinect. The RGB color information is analyzed as shown in Fig. 4(a), (b), and the depth values are computed from Eqs. (10)–(12). Table I shows a data set for the relationship between color and the corresponding distance values, and the measurements and RGB color estimation depth values are compared in Fig. 5. RGB-D cameras are powerful tools for robot perception and help detect projected patterns based on surface reflections. Vital advancements in visual robotics consist of three core technologies: visual pattern recognition (ViPR), V-SLAM, and optical analysis

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

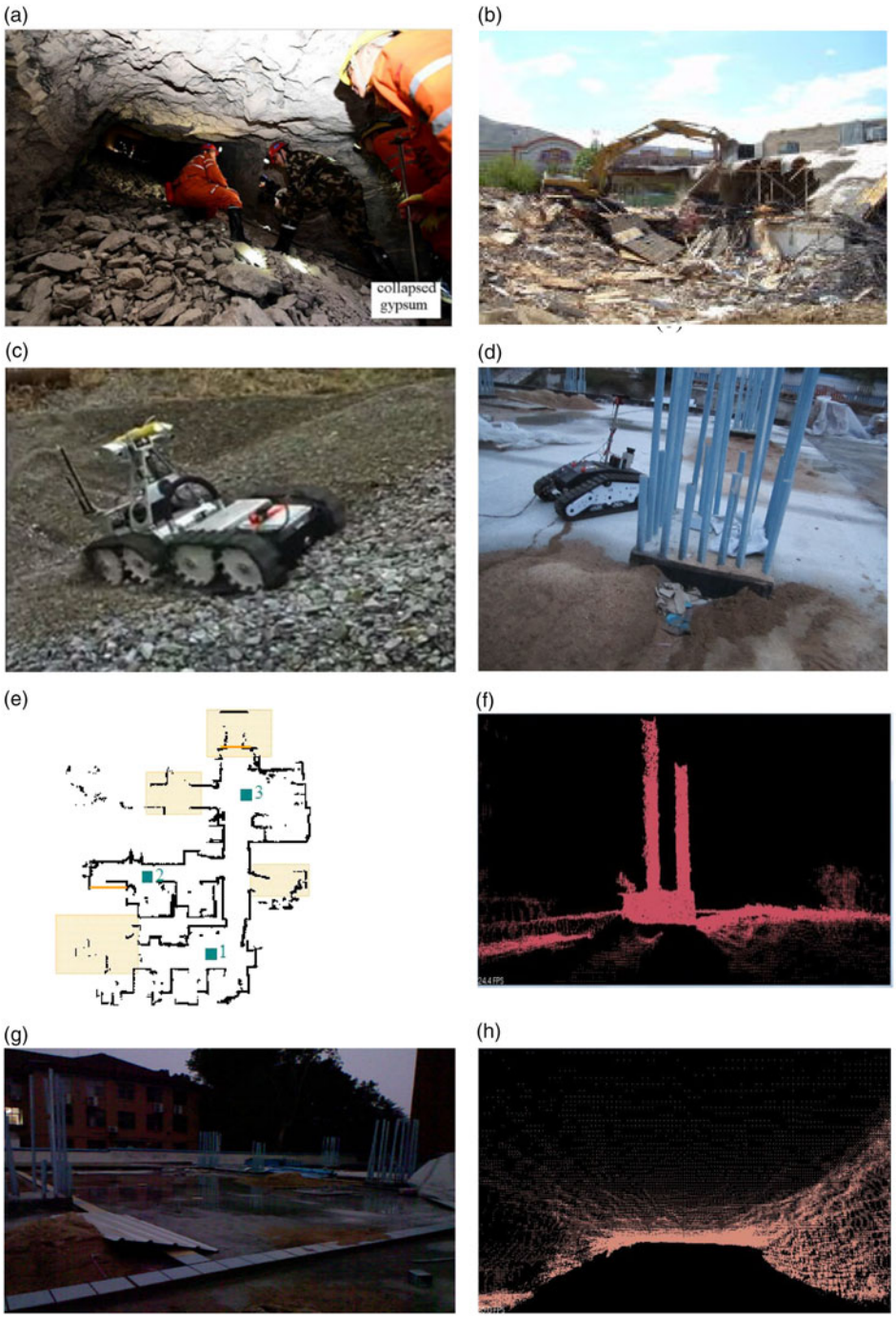

Fig. 14. A LUKER joint-type crawler mobile robot explores a collapsed scenario. (a) A SAR team in the pose-disaster scene. (b) A demolishment collapse environment. (c) A crawler mobile robot performing SAR tasks. (d) A robot equipped with Kinect RGB-D camera and LRFs explores at a collapsed site. (e) The generated 2D map. (f) The 3D reconstruction depth map. (g) The robot obtaining the collapse site scene. (h) The reconstructed depth map.

for localization. 3D visual reconstruction SLAM is a procedure for V-SLAM using RGB-D camera sensors combined with dead-reckoning odometry information. Figures 1(a), 3 and 6(a) show RGB-D camera systems that are used to obtain image frames of OOIs for robot SLAM adapted to SAR environments. Table I and Fig. 5 analyze the errors by comparing the actual distance with RGB-D depth estimation in 3D reconstruction in the 2D SLAM process. From Table I, the estimation depth agrees well with the actual distance in 3D reconstruction SLAM. Figure 7(a), (b) test the 3D identification

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

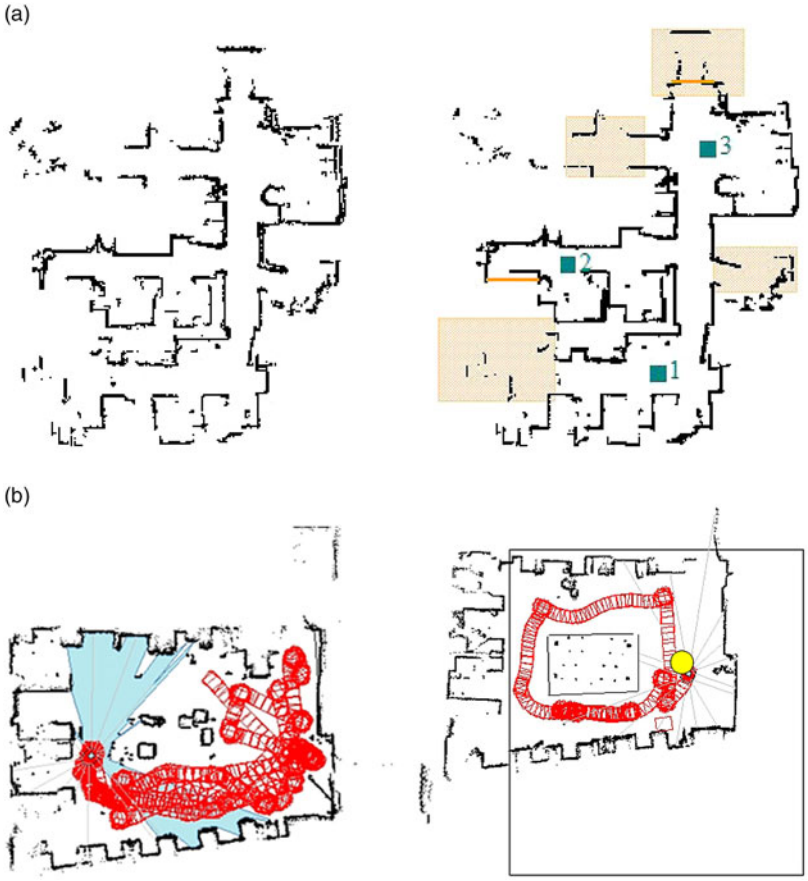

Fig. 15. Comparison of 2D/3D reconstruction SLAM with a single 2D map. (a) Comparing with Fig. 14(e) regarding OI. (b) Comparing with Fig. 12(e) regarding loop-closure detection.

Algorithm 2 according to Eq. (13) and the program of Table II, which highlights the colored OOIs (i.e., the red outlined areas).

Figures 11, 12, and 14 show several experimental scenarios. Using the IPC that is shown in Fig. 12(a)–(f), the robot's SLAM process is monitored by remote PC control center. Figure 14(f)–(h) show 3D reconstructions of the exploration environment with the installed RGB-D Kinect and the localized sensors; the depth information about the OOIs in the collapsed SAR environment and 3D reconstruction are based on the 2D map. The fast and simple 2D map performances improved the 3D OI in the 2D SAR SLAM methods.

7. Conclusions and Future Work

In this paper, the OI method of 3D local target reconstruction based on 2D SLAM is studied. Compared with 3D V-SLAM, the proposed method maintains the rapidity and real-time performance of 2D map construction. Compared with 2D SLAM, the proposed method can create SAR-object 3D local scenes and detect LC point revisits. The main novel contributions of this paper are as follows.

First, a theoretical model is established that is adapted to the SAR environment, and models are developed for OI and detecting LC point revisits. By combining the advantages of RGB-D image processing, IICP, and other algorithms, a 2D/3D fusion method is proposed for creating the SAR environment map. The mobile robot relies on planar distance estimation to construct a 2D map, and the coverage of SLAM signal exploration area is detected. The 3D object local map is then reconstructed using the proposed mathematical models of RGB-ICP and 2D/3D wavelet transform, both of which are shown in Algorithm 3 and Eqs. (19)–(21).

Second, a method based on the J–G[17] mathematical model and an OpenCV processing algorithm (http://opencv.org/) is proposed to solve OI and LC problem is proposed. For the first time, the mathematical theory of optical flow calculation[12] is used to solve robotic SAR SLAM problem, and

https://doi.org/10.1017/S0263574719000675 Published online by Cambridge University Press

the new contributions of this paper are detailed in Eqs. (9)–(18), Tables I and II, and Algorithm 2. Theoretical analysis shows that this algorithm enables mobile robots to conduct SAR environmental SLAM autonomously, as shown in Eqs. (1)–(8) and Algorithm 1.

Third, because Visual Studio, Matlab, MobileRobot, and ROS are all common software packages used for conducting robotic SLAM, we can used them for mobile robots running the SAR SLAM process autonomously.

We used different mobile robots to conduct SAR SLAM to provide strong evidence for the innovation contributions in this paper. The 2D/3D SLAM and the SAR OOIs detection were implemented in simulations and robot ground-truth experiments, as shown in Figs. 8 and 10–14.

The designed V-SLAM RGB-D systems provide image depth information of OOIs in unstructured SAR environments. The Kinect camera and LRF sensors are used to obtain the depth distance between the mobile robot and the OOIs within the SAR environment. Using the proposed algorithms and image processing methods, the SLAM performances of LC and OI in harsh SAR environments. The experiments with the crawler robot validated (i) the efficacy of outdoor SLAM and (ii) the proposed 3D objects reconstruction V-SLAM on basis of 2D mapping methods. This process includes three main phases: (i) the spatial alignment of consecutive image frames (3D objects identification), (ii) the detection of LCs (SAR area autonomous exploration), and (iii) globally consistent matching of the complete data sequence (global 2D map building). The robot's pose is represented by 3D position and orientation data, and the OOIs within SAR environment are mapped with natural landmarks of the metric map in 3D reconstruction. A robot communicates with other robots onboard RGB-D, Kinect multisensors, environmental signals, and remote PC control center, which may supply adequate depth information. The communication and data exchange among several multi-robot systems and the remote PC control center requires a robust network. Given that GPS signal can be either unavailable or unreliable in a SAR collapse environment, in another research work, we have developed an *ad hoc* wireless local network for information exchange among multiple robots, multisensors, and the remote PC control center. In future work, we plan to develop the fusion of internet and GPS technologies to supplement rich depth information. Additionally, the optical flow analysis computation algorithms can be refined for 3D visual reconstruction by extracting RGB image depth values.

Acknowledgments

This work is supported by the NSFC (National Nature Science Foundation of China) under grant nos. 61573213 and 61673245; by the National Key Research and Development Plan of China under grant no. 2017YFB1300205; and by Shandong Province Key Research and Development Plan under grant no. 2016ZDJS02A07.

References

1. G. Y. Lin and Y. T. Wang, "Improvement of speeded-up robust features for robot visual simultaneous localization and mapping," *Robotica* **32**(2), 533–549 (2014).
2. H. Martins, I. Oakley and R. Ventura, "Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots," *Robotica* **33**(10), 2166–2185 (2015).
3. R. R. Murphy, "Trial by fire," *IEEE Rob. Autom. Mag*. **11**(9), 50–61 (2004).
4. Y. Yokokohji, M. Kurisu, S. Takao, et al., "Constructing a 3D Map of Rubble by Teleoperated Mobile Robots with a Motion Canceling Camera System," *Proceedings of the 2003 IEEE/RSJ, International Conference on Intelligent Robots and Systems*, Las Vegas, Nevada (2003) pp. 3118–3125.
5. R. R. Murphy, J. Kravitz, S. L. Stover and R. Shoureshi, "Mobile robots in mine rescue and recovery," *IEEE Rob. Autom. Mag*. **9**(6), 91–103 (2009).
6. J. M. Santos, D. Portugal and R. P. Rocha, "An Evaluation of 2D SLAM Techniques Available in Robot Operating System," *Fundaçào para a Ciência e a Tecnologia*, The Portuguese Science Agency (2013).
7. L. Alboul and G. Chliveros, "A System for Reconstruction from Point Clouds in 3D: Simplification and Mesh Representation," *2010 11th International Conference on Control, Automation, Robotics and Vision*, Singapore (2010) pp. 2301–2306.
8. P. Henry, M. Krainin, E. Herbst, X. Ren and D. Fox, "RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments," *Int. J. Rob. Res*. **31**(5), 647–663 (2012).
9. S. Weiss, D. Scaramuzza and R. Seigwart, "Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments," *J. Field Rob*. **28**(6), 854–874 (2011).
10. K. Nagatani, H. Ishida, S. Yamanaka and Y. Tanaka, "Three-Dimensional Localization and Mapping for Mobile Robot in Disaster Environments," *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, Nevada (2003) pp. 3112–3117.

11. M. McGill, R. Selleh, T. Wiley, et al., "Virtual Reconstruction using an Autonomous Robot," *2012 International Conference on Indoor Positioning and Indoor Navigation*, Sydney, Australia (2012).
12. M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Machine Vision*, 3rd. ed. (Thomson Corporation, Toronto, 2008). http://www.thomsonlearning.com.
13. J. V. Miró, W. Z. Zhou and G. Dissanayake, "A Strategy for Efficient Observation Pruning in Multi-objective 3D SLAM," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, CA, USA (2011) pp. 1640–1646.
14. Z. Riaz, T. Linder, S. Behnke, R. Worst and H. Surmann, "Efficient Transmission and Rendering of RGB-D Views," *Advances in Visual Computing*, **In:** *ISVC, Part I, LNSC* (G. Bebis et al. eds.), vol. 8033 Springer Berlin Heidelberg (2013) pp. 517–26.
15. S. Fazli and L. Kleeman, "Simultaneous landmark classification, localization and map building for an advanced sonar ring," *Robotica* **25**(3), 283–296 (2007).
16. M. J. Veth, R. K. Martin and M. Pachter, "Anti-temporal-aliasing constraints for image-based feature tracking applications with and without inertial aiding," *IEEE Trans. Veh. Technol*. **59**(8), 3744–3756 (2010).
17. Wikipedia, the free encyclopedia, "HSL and HSV," (2015). http://en.wikipedia.org/wiki/HSL_and_HSV.
18. W. Z. Zhou, J. V. Miró and G. Dissanayake, "Information-efficient 3-D visual SLAM for unstructured domains," *IEEE Trans. Rob*. **24**(5), 1078–1087 (2008).
19. A. Nüchter, K. Lingemann, J. Hertzberg and H. Surmann, "6D SLAM - 3D mapping outdoor environments," *Fraunhofer Institute for Autonomous Intelligent Systems (AIS) Schloss Birlinghoven D-53754*, Sankt Augustin, Germany (2007).
20. G. Nejat and Z. Zhang, "Finding disaster victims: robot-assisted 3D mapping of urban search and rescue environments via landmark identification," *IEEE ICARCV*. **1**(6), 50–61 (2006).
21. Z. Zhang, H. Guo, G. Nejat and P. Huang, "Finding Disaster Victims: A Sensory System for Robot-Assisted 3D Mapping," *IEEE International Conference on Robotics and Automation*, Rome, Italy (2007) pp. 3889–3894.
22. Z. Zhang and G. Nejat, "Robot-Assisted Intelligent 3D Mapping of Unknown Cluttered Search and Rescue Environments," *IEEE/RSJ International Conference on Intelligent Robots and System*, Acropolis Convention Center, Nice, France (2008) pp. 2115–2120.
23. M. Liu, F. Colas, L. Oth and R. Siegwart, "Incremental topological segmentation for semi-structured environments using discretized GVG," *Auton. Rob*. **38**(2), 143–160 (2015).
24. R. S. Blum and Z. Liu, *Multi-Sensor Image Fusion and Its Applications* (CRC Press, Taylor & Francis Group, Oxford, 2006).
25. D. Schleicher, L. M. Bergasa, M. Ocaña, R. Barea and M. E. López, "Real-time hierarchical outdoor SLAM based on stereovision and GPS fusion," *IEEE Trans. Intell. Transp. Syst*. **10**(3), 50–61 (2009).
26. I. Bloch, *Information Fusion in Signal and Image Processing* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008).
27. A. V. Savkin and M. Hoy, "Reactive and the shortest path navigation of a wheeled mobile robot in cluttered environments," *Robotica* **31**(2), 323–330 (2013).
28. F. Aghili, "3D simultaneous localization and mapping using IMU and its observability analysis," *Robotica* **29**(10), 805–814 (2011).
29. T. Deißler and J. Thielecke, "UWB SLAM with Rao-Blackwellized Monte Carlo Data Association," *International Conference on Indoor Navigation (IPIN)*, Zürich, Switzerland (2010).
30. M. Happold and M. Ollis, "Using learned features from 3D data for robot navigation," *Stud. Comput. Intell. (SCI)* **76**, 61–69, Applied Perception, Inc., Cranberry Township, Pennsylvania (2007).
31. H. Surmann, A. Nüchter and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," *Rob. Auton. Syst*. **45**(9), 181–198 (2003).
32. J. Bêdkowski, K. Majek, P. Musialik, A. Adamek, D. Andrzejewski and D. Czekaj, "Towards terrestrial 3D data registration improved by parallel programming and evaluated with geodetic precision," *Autom. Constr*. **47**(8), 78–91 (2014).
33. K. Ohno, T. Nomura and S. Tadokoro, "Real-Time Robot Trajectory Estimation and 3D Map Construction Using 3D Camera," *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (2006) pp. 5279–5285.
34. T. Bücher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage and W. von Seelen, "Image processing and behavior planning for intelligent vehicles," *IEEE Trans. Ind. Electron*. **50**(1), 62–75 (2003).
35. L. P. Ellekilde, S. D. Huang, J. V. Miró and G. Dissanayake, "Dense 3D map construction for indoor search and rescue," *J. Field Rob*. **24**(1–2), 71–89 (2007).
36. F. Jesus and R. Ventura, "Combining monocular and stereo vision in 6D-SLAM for the localization of a tracked wheel robot," *IEEE Inst. Syst. Rob*. **12**(4), 50–61 (2012).
37. R. H. Wong, J. Z. Xiao and S. L. Joseph, "An Adaptive Data Association for Robotic SLAM in Search and Rescue Operation," *Proceedings of the 2011 IEEE International Conference on Mechatronics and Automation*, Beijing, China (2011) pp. 997–1003.
38. J. Weingarten and R. Siegwar, "3D SLAM Using Planar Segments," *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (2006) pp. 3062–3067.
39. A. Carbone, D. Ciacelli, A. Finzi and F. Pirri, "Autonomous Attentive Exploration in Search and Rescue Scenarios," **In:** *WAPCV* (L. Paletta and E. Rome, eds.), (2007) pp. 431–446. DOI: 10.1007/978-3-540-77343-6_28.

40. D. Schleicher, L. M. Bergasa, M. Ocaña, R. Barea and E. López, "Real-time hierarchical stereo visual SLAM in large-scale environments," *Rob. Auton. Syst*. **58**(8), 991–1002 (2010).
41. W. Z. Zhou, J. V. Miró and G. Dissanayake, "Information-Driven 6D SLAM Based on Ranging Vision," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Acropolis Convention Center, Nice, France (2008) pp. 2072–2077.
42. D. Stronger and P. Stone, "Selective Visual Attention for Object Detection on a Legged Robot," **In:** *RoboCup, LANI* (Lakemeyer et al. eds.), vol. 4434 Springer-Verlag (2007) pp. 158–170.
43. E. Mihankhah, H. D. Taghirad, A. Kalantari, E. Aboosaeedan and H. Semsarilar, "Line Matching Localization and Map Building with Least Square," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Suntec Convention and Exhibition Center*, Singapore (2009) pp. 1734–1739.
44. A. Kim and R. M. Eustice, "Perception-Driven Navigation: Active Visual SLAM for Robotic Area Coverage," *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany (2013) pp. 3196–3203.
45. K. P. Valavanis, L. Doitsidis, M. Long and R. R. Murphy, "A case study of fuzzy-logic-based robot navigation," *IEEE Rob. Autom. Mag*. **6**(9), 93–107 (2006).
46. B. Day, C. Bethel, R. Murphy and J. Burke, "A Depth Sensing Display for Bomb Disposal Robots," *Proceedings of the 2008 IEEE International Workshop on Safety, Security and Rescue Robotics*, Japan (2008) pp. 146–151.
47. Z. Zhang and G. Nejat, "Intelligent sensing systems for rescue robots: landmark identification and three-dimensional mapping of unknown cluttered urban search and rescue environments," *Adv. Rob*. **23**(11), 1159–1177 (2009).
48. Y. Yamamoto, P. Pirjanian, M. Munich, E. DiBernardo, L. Goncalves, J. Ostrowski and N. Karlsson, "Optical Sensing for Robot Perception and Localization," *IEEE Workshop on Advanced Robotics and its Social Impacts*, Nagoya, Japan (2005) pp. 14–17.
49. T. Fujiwara, T. Kamegawa and A. Gofuku, "Stereoscopic Presentation of 3D Scan Data Obtained by Mobile Robot," *Proceedings of the 2011 IEEE International Symposium on Safety, Security and Rescue Robotics*, Kyoto, Japan (2011) pp. 178–183.
50. J. Knuth and P. Barooah, "Distributed collaborative 3D pose estimation of robots from heterogeneous relative measurements: an optimization on manifold approach," *Robotica* **33**(7), 1507–1535 (2014).
51. T. Pire, R. Baravalle, A. D'Alessandro and J. Civera, "Real-time dense map fusion for stereo SLAM," *Robotica* **36**(10), 1510–1526 (2018). doi:10.1017/S0263574718000528
52. M. R. U. Saputra, A. Markham and N. Trigoni. "Visual SLAM and structure from motion in dynamic environments: a survey," *ACM Comput. Surv*. **51**(2), 1–37 (2018).