The UMA-SAR Dataset: Multimodal Data Collection from a Ground Vehicle during Outdoor Disaster Response Training Exercises

This is the ACCEPTED VERSION of an article published in:
The International Journal of Robotics Research. 2021;40(6-7):835-847.
https://doi.org/10.1177/02783649211004959

Jesús Morales, Ricardo Vázquez-Martín, Anthony Mandow, David Morilla-Cabello and Alfonso García-Cerezo

Abstract

This article presents a collection of multimodal raw data captured from a manned all-terrain vehicle in the course of two realistic outdoor search and rescue (SAR) exercises for actual emergency responders conducted in Málaga (Spain) in 2018 and 2019: the UMA-SAR dataset. The sensor suite, applicable to unmanned ground vehicles (UGV), consisted of overlapping visible light (RGB) and thermal infrared (TIR) forward-looking monocular cameras, a Velodyne HDL-32 three-dimensional (3D) lidar, as well as an inertial measurement unit (IMU) and two global positioning system (GPS) receivers as ground truth. Our mission was to collect a wide range of data from the SAR domain, including persons, vehicles, debris, and SAR activity on unstructured terrain. In particular, four data sequences were collected following closed loop routes during the exercises, with a total path length of 5.2 km and a total time of 77 min. Additionally, we provide three more sequences of the empty site for comparison purposes (an extra 4.9 km and 46 min). Furthermore, the data is offered both in human-readable format and as rosbag files, and two specific software tools are provided for extracting and adapting this dataset to the users' preference. The review of previously published disaster robotics repositories indicates that this dataset can contribute to fill a gap regarding visual and thermal datasets and can serve as a research tool for cross-cutting areas such as multispectral image fusion, machine learning for scene understanding, person and object detection, and localization and mapping in unstructured environments. The full dataset is publicly available: www.uma.es/robotics-and-mechatronics/sar-datasets.

Keywords

Disaster robotics, Search and rescue, multimodal sensors, multispectral imaging, thermal infrared camera, 3D lidar, dataset

1 Introduction

Disaster robot missions such as reconnaissance and mapping, search for survivors, logistics, first medical assistance, casualty evacuation, and support for cooperative perception could benefit from research in new artificial intelligence solutions for multimodal perception. A single search and rescue (SAR) robot can produce gigabytes of heterogeneous types of imagery and data (Murphy et al. 2016), so new intelligent tools are required to prevent information overload for end-users. Besides, perception in disaster situations poses distinctive challenges due to limited image quality, a diversity of target objects, unstructured search environments and integration of visual information with other data modalities (Arnold et al. 2019). Thus, the availability of representative datasets becomes essential for developing, training, testing and benchmarking new artificial intelligence solutions for perception, classification, localization and mapping in the challenging disaster response application domain (Petříček et al. 2019)(Bañuls et al. 2020)(Dubé et al. 2020).

This article presents a public sensor-oriented dataset obtained from a manned all-terrain vehicle during realistic disaster response exercises on outdoor terrain (see Figure 1). Our mission was not to participate in a particular rescue scenario but to capture a wide range of data from the SAR domain. The vehicle was equipped with a specific multimodal sensor suite to collect data appropriate for machine learning object recognition, disaster site modeling, and situational awareness research in unmanned ground vehicles (UGV). In particular, the sensor suite incorporated a combination of overlapping thermal infrared (TIR) and visible light (RGB) cameras, three-dimensional (3D) lidar, an inertial measurement unit (IMU) and global positioning system (GPS) receivers.

The main aspects of our dataset are the following:

- The data has been acquired during large-scale exercises on unstructured terrain organized for actual responders. The realistic mockup disaster scenarios (physical fidelity) and the behavior of the rescuers and other participants (personnel fidelity) were not specifically staged for this work.
- Each data sequence closes a path loop with GPS ground truth for SLAM research. Data has been captured from a manned vehicle for ensuring safety and maneuverability.

Universidad de Málaga, Andalucía Tech, Robotics and Mechatronics Group, 29071 Málaga, Spain.

Corresponding author:
Anthony Mandow. Email:amandow@uma.es

Figure 1. Photographs of the site (top) and two rescue scenarios (middle, bottom).

- We offer overlapping TIR and RGB image pairs with a wide range of SAR related object categories for multi-spectral machine learning. These include numerous instances of persons (e.g., rescuers, observers, and simulated victims), vehicles (e.g., rescue and military vehicles, cars, and vans), debris (e.g., rubble, sewer pipes, and crushed cars) and SAR activity.
- Our data is presented in human readable format along with software tools for adjusting and selecting the data to the preference and workflow of each user.

Our major purpose is that the research community can benefit from our opportunity to collect ground vehicle data from outdoor field events designed with substantial authenticity by and for actual SAR professionals. The remaining of this data paper is organized as follows. The next section reviews published datasets for disaster robotics. Section 3 presents the disaster exercises and site, the vehicle-sensor setup, and the data collection methodology. Section 4 describes the dataset and the accompanying software tools. Section 5 discusses limitations of the data as well as some of the ways in which the dataset could be used. Finally, Section 6 offers a summary.

2 Related Work

Table 1 summarizes published datasets for robotics in the SAR domain in chronological publication order. Web links are provided for the datasets that are publicly available at the time of this writing. In the cases where original links are lost, these might be reached by request. Besides, the table indicates the viewpoint (aerial or ground sensors), a brief description of the site, the major data modalities, the main purpose of the dataset, and the presence of human-figure subjects in the captured data. It can be observed that the major purpose for most datasets was related either to site modeling and SLAM or to machine learning.

The major motivations for the first datasets were simultaneous localization and mapping (SLAM) and 3D modeling of disaster environments, which are fundamental capabilities for SAR UGVs (Droeschel et al. 2017)(Dubé et al. 2020). Several works used Disaster City (Texas), a training facility with a variety of realistic mock-up SAR scenarios (e.g., a collapsed parking building and a train accident), to collect UGV-based datasets with physical fidelity, such as 3D lidar scans for mapping (Ohno et al. 2010) and combinations of lidar and RGB images for terrain classification and SLAM (Pellenz et al. 2010) and 3D representations with a simulated victim (Birk et al. 2009)(Pathak et al. 2010). This facility was also used to test the RESPOND-R data management framework during exercises with a UAV-UGV team, where data from heterogeneous sources, such as communications or human-robot interactions, were collected in addition to video and scans (Shrewsbury et al. 2013)(Duncan and Murphy 2014). On the other hand, the repositories of the disaster robotics research projects ICARUS (Balta et al. 2017) and TRADR (Svoboda 2017) offer datasets collected from unmanned ground (UGV) and aerial vehicles (UAV). The ICARUS datasets consist of point clouds captured at a military base used for SAR training, with the major goal of producing 3D reconstructions of buildings and disaster sites. These point clouds were derived mostly from different UAV cameras (either RGB, grayscale or TIR) with GPS annotations, but also from a UGV with a high-resolution 3D lidar. The TRADR dataset compilation includes proprioceptive readings and a variety of sensor setups for different robotic applications, such as omnidirectional cameras and lidar for stair climbing with tracked robots (Kubelka et al. 2019) and multi-robot point cloud registration in disaster sites (Gawel et al. 2017)(Dubé et al. 2017). Even if the primary purpose of the TRADR datasets was documenting individual methods and experiments, the repository shares data from realistic sites such as decommissioned plants, outdoor environments and firefighting training facilities. Furthermore, the TEDUSAR dataset Leingartner et al. (2016) resulted from a relevant effort to capture realistic multimodal data from a full-scale exercise for actual responders in a GPS-denied tunnel accident scenario with vehicles and hazardous materials. The major mission of the TEDUSAR UGV was collecting data to be shared with the research community for assessing performance of a variety of state-of-the-art sensors and mapping algorithms. Originally, the sensor suite comprised a 3D lidar, TIR, stereo, omnidirectional and Kinect cameras, an IMU, and an arm with a 2D lidar and a radar sensor, but eventually only the RGB cameras and lidar worked on account of the challenging conditions and timing constraints imposed by the emergency response drill. More recently, the HRAIL repository at Texas A&I University (HRAIL 2020), holds a collection of aerial image datasets captured in the aftermath of actual natural disasters, which are valuable for mapping, debris/building inspection, and property damage assessment.

In the last years, the emergence of machine learning as an effective research tool for intelligent visual perception

Table 1. Published datasets in the SAR domain (in chronological order).

Dataset	Availability	Viewpoint	Site	Modalities	Main purpose	Humans
Jacobs University (Birk et al. 2009) (Pathak et al. 2010)	Public[1]	UGV	Disaster city (collapsed building)	3D lidar	SLAM	Victim stand-in
Koblenz-Landau University (Pellenz et al. 2010)	Public[2]	UGV	Disaster city (outdoor)	3D lidar, RGB, GPS	Terrain classification and mapping	-
Tohoku University (Ohno et al. 2010)	By request	UGV	Disaster city (indoor and outdoor)	3D lidar	SLAM	-
RESPOND-R (Shrewsbury et al. 2013)	Authorised users	UAV, UGV, ground	Disaster city (outdoor)	RGB video, state, GPS, environmental	Radiological localization and assessment. Data logging	-
TEDUSAR (Leingartner et al. 2016)	By request	UGV	Tunnel exercise for rescuers	RGB, 3D lidar.	SLAM. Sensor evaluation	-
ICARUS (Balta et al. 2017)	Public[3]	UAVs and one UGV	Military base (outdoor, indoor)	Point clouds from RGB, TIR or monochrome cameras (UAVs) and 3D lidar (UGV)	Data management and site modeling for rescuers	-
TRADR Compilation (Svoboda 2017)	Public[4]	Mostly UGV. UAV	Several Indoor / outdoor	Mostly RGB with robot status. 3D lidar	SLAM, exploration and navigation	-
RoboCup Rescue Victim (Lorenz and Steinbauer 2018)	Public[5]	UGV	RoboCup Rescue arena (indoor)	Annotated RGB	Victim face detection. Machine learning	Doll baby victims
DISC (Jeon et al. 2019)	Public[6]	Ground	Virtual indoor and outdoor	Stereo with ground truth	Machine learning. Structure damage	-
TRADR Active segmentation (Petříček et al. 2019)	Public[7]	UGV	Indoor mockup disasters	RGB-D and TIR, state	Machine learning. Victim segmentation. Camera control	Semi-synthetic victims
Leading India SAR (Mishra et al. 2020)	By request	UAV	Campus (outdoor)	RGB	Machine learning. Gestures	Pedestrians
SDV (Dadwhal et al. 2020)	Public[8]	Ground	Cluttered Indoor	RGB + ground truth	Machine learning. Skin detection	Victim stand-ins
HRAIL Texas A&I University (HRAIL 2020)	Authorised users	UAV	Actual disasters (outdoor)	RGB	Mapping, damage assessment	-

[1] http://robotics.jacobs-university.de/media
[2] http://kos.informatik.uni-osnabrueck.de/3Dscans
[3] http://projects.asl.ethz.ch/datasets/doku.php?id=jfricarus
[4] http://www.tradr-project.eu/resources/datasets
[5] https://osf.io/dwsnm
[6] https://sites.google.com/site/hgjeoncv/disc-project-page
[7] http://ptak.felk.cvut.cz/tradr/data/human_seg
[8] http://dx.doi.org/10.21227/a5dm-y470

has demanded image datasets with representative domain-specific objects for model training. In some cases, the lack of suitable datasets can be partially mitigated by computing synthetic images, as in the DISC dataset (Jeon et al. 2019), which consists of sufficiently realistic stereo renderings of fire and collapsing structures that allow comparing before and after disaster conditions. In this sense, Petříček et al. (2019) created a semisynthetic dataset consisting of chroma key images of human victim stand-ins that can be overlaid onto background images of disaster environments. Precisely, recent datasets have incorporated images of human figures for intelligent victim detection. Such is the case of the RGB dataset collected by Lorenz and Steinbauer (2018) during the 2015 RoboCup Rescue competition, where doll babies simulated victims in a cluttered indoor arena. Similarly, Dadwhal et al. (2020) staged the simulated disaster victim (SDV) datasets for skin detection in images of dust-covered humans lying among indoor rubble. In contrast, the purpose of the LeadingIndia-SAR dataset (Mishra et al. 2020) was recognizing help requests and other human actions from aerial images in an unmodified campus environment. In addition to visible light images, thermal images can be useful for scene understanding in texture-less, dark, and smoke- or dust-filled SAR settings. Petříček et al. (2019) addressed the combination of thermal and visual images for multimodal victim detection from a robotic vehicle in mock-up disaster sites. Their dataset documents the research goal of controlling the pan-and-tilt of a thermal camera for human body segmentation of a 3D occupancy grid built from motion estimations, panoramic RGB images and depth information. Apart from that, existing thermal image datasets do not address SAR but other applications like nighttime pedestrian detection (Xu et al. 2019) and surveillance Krišto and Ivašić-Kos (2019), where performance can be improved with multispectral combinations of color and thermal cameras, as in the KAIST dataset (Choi et al. 2018).

The motivations for our dataset are similar to those of the TEDUSAR dataset (Leingartner et al. 2016) in the sense that our primary goal has been sharing multimodal data collected from a ground vehicle in the context of full-scale training exercises for response personnel. However, our data (which includes RGB, TIR and 3D lidar) has been obtained from different types of outdoor scenarios and we have been able to include GPS readings for ground truth. Furthermore,

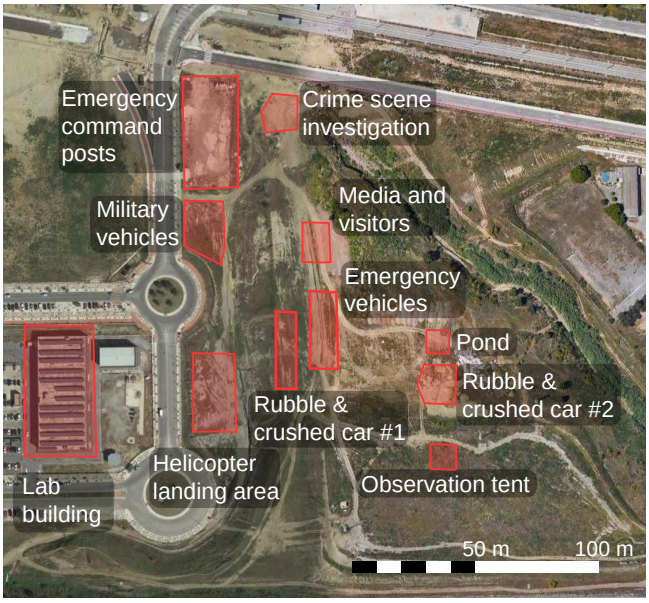

Figure 2. Aerial view of the disaster simulation site (Google 2020), where shaded red areas indicate objects and scenarios recognizable in the dataset.

our dataset contributes to filling the observed gap regarding thermal and visual image datasets in the disaster response domain, as it includes synchronized overlapping images of both spectra with similar fields of view (FOV). Besides, our multimodal data not only captures closed loops of terrain information for 3D modeling and SLAM but also different types of human subjects (e.g., responders, pedestrians, victims) and other SAR-related objects and actions, which can be useful for machine learning.

3 Data collection

3.1 Disaster exercises

The UMA-SAR dataset was captured during two realistic disaster response exercises that took place in Málaga (Spain) on June 1, 2018 and on June 6, 2019, respectively. The exercises were conducted as part of an annual Workshop organized by the Chair of Security, Emergencies and Disasters at Universidad de Málaga (UMA). More than one hundred registered personnel from different governmental and non-governmental organizations participated in each exercise.

Each full-scale exercise involved a series of scenarios corresponding to the response to a crisis situation, namely an earthquake in 2018 and a terrorist attack in 2019. The exercise site was a dedicated $90{,}000\,\mathrm{m}^2$ outdoor experimental area within the UMA campus. This natural terrain area was set up as a simulated disaster site, including rubble mounds, crushed vehicles, and partially buried sewer pipes, as illustrated in Figure 1 (top).

A layout of the site with the main areas and activities captured in the dataset is depicted in Figure 2. During the exercises, SAR units from participant organizations deployed their tents, vehicles and personnel in an area for *emergency command posts*. In the 2018 data capture, the rescue scenarios were prepared in the *rubble and crushed car #1* area, the *rubble and crushed car #2* area, and the *pond*.

Figure 3. All-terrain Argo vehicle with the sensor suite on the front rack. The photograph was taken during the 2019 exercise.

Besides, *emergency vehicles* as well as numerous registered participants, including rescuers, a canine unit, and *media and visitors*, were standing or preparing their equipment on the sides of the dirt trail that connected the *emergency command posts* with the rescue scenarios. In 2019, a rescue scenario was conducted in the *rubble and crushed car #2* area, with victims inside the car, and the aftermath of an attack was set up at the *crime scene investigation* area.

The exercises were conducted under planned strict timing constraints and safety protocols coordinated by the exercise director and the organization staff. This complex coordination effort involved, for instance, air traffic control constraints. Other personnel participating in the exercise included the first response teams, victim stand-ins (played by drama students), authorized media, visitors and officials. Some of the scenarios were used to test robotic technologies (Fernández-Lozano et al. 2018) but these were not related to our dataset.

3.2 Vehicle mission and sensor setup

Our mission in the exercises, which was clearly defined in advance and approved by the organizers, was to deploy a manned all-terrain vehicle within the exercise site to collect multimodal imaging and data from the developing SAR scenarios. This data collection mission was not limited to active scenarios but also to the transit between the different areas within the exercise site, which also included representative objects such as personnel and vehicles. Thus, our mission was independent of other missions conducted by responders and we were expected to minimize interference with their activity. Furthermore, the main object of data capture was to serve as a research tool for both SLAM and machine learning in disaster robotics and other related technologies, so it was not meant to be used by emergency response teams during or after the exercises.

The vehicle used for data capture was an 8×8 Conquest 1050 XTD by Argo (see Figure 3) that allows skid-steer maneuverability through rough and uneven terrain. The sensor suite platform was designed to fit onto the vehicle's front rack, which is $1.2\,\mathrm{m}$ above and parallel to the ground plane. This platform also accommodated the data acquisition devices, i.e., communication devices and a computer.

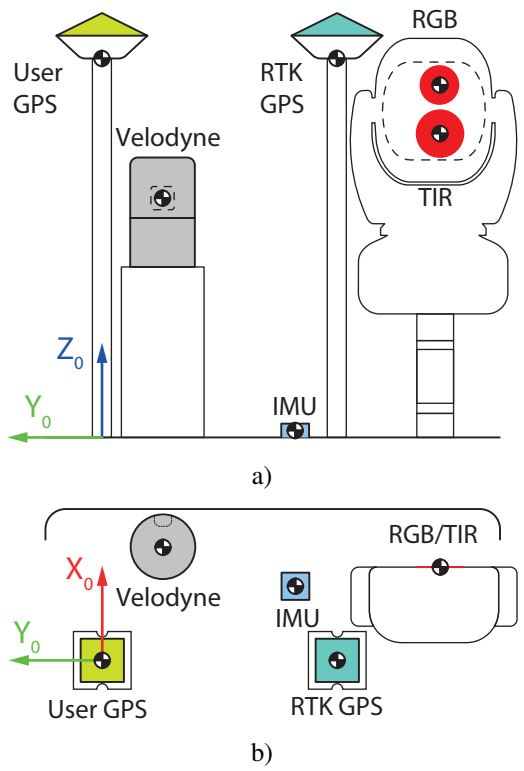

Figure 4. Sensor suite layout: a) Back view (i.e., driver standpoint), and b) top view. The center of mass symbol indicates the origin of each sensor's frame.

In particular, the following sensors were included (see Figure 4):

- **RGB and TIR cameras.** We used an Oculus-TI camera unit (by Silent Sentinel) with monocular TIR and RGB analog cameras mounted on a pan-tilt-zoom (PTZ) system. Both images are mostly overlapped, allowing image fusion research. The microbolometer produces thermal images in the longwave infrared (LWIR) wavelength band (i.e., 8 to 14μm), which were recorded with white hot polarity. The PTZ was fixed to forward looking orientation with the maximum horizontal FOV, which is 44° and 57.8° for TIR and RGB, respectively. An AXIS P7214 IP video server captures and digitizes images from both cameras with a resolution of 704×576 at 25 Hz.
- **3D lidar.** We used the Velodyne HDL-32, with a vertical 32-beam FOV of 41.33° (from −30.67° to 10.67°) and a measurement range of up to 100 m. A scan frame consists on a full rotation in the 360° horizontal FOV. The point clouds were acquired at 10 Hz, which gives a vertical angular resolution of 1.33° and an azimuthal resolution of 0.166°.
- **Global positioning system real-time kinematics (GPS-RTK).** We included two JAVAD L1-band GPS receivers with MarAnt+ antennas to gather position information at 5 Hz. One was configured in standalone mode (user) and the other in fixed mode (RTK) for centimeter accuracy. In 2018, we placed our own JAVAD base station with an UHF Radio link at the exercise site. In 2019, the corrections were taken from the base station MLGA 13460M001 of the regional public positioning network (Berrocoso et al. 2006) through 4G internet connection. This base station is close to the exercise site (approximately 5 km) and streams corrections with a Networked Transport of RTCM via Internet Protocol (NTRIP) server.
- **IMU**. We used Microstrain IMUs to measure linear acceleration, angular velocity and orientation. In 2018, the model was a 3DM-GX2 with a nominal capture rate of 100 Hz and in 2019, it was a 3DM-GX5-25 at 64 Hz.

Table 2. Sensor frame definitions with respect to the vehicle's frame for the 2018 setup.

	x (cm)	y (cm)	z (cm)	$\vec{x}$	$\vec{y}$	$\vec{z}$
User GPS	0	0	50			
RTK GPS	0	−18	50			
IMU	16.9	−40.9	1.6	$-\vec{y}_0$	$\vec{x}_0$	$-\vec{z}_0$
Velodyne	21.3	−9.3	31.6	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$
RGB	7.4	−40.9	46.5	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$
TIR	7.4	−40.9	40	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$

Table 3. Sensor frame definitions with respect to the vehicle's frame for the 2019 setup.

	x (cm)	y (cm)	z (cm)	$\vec{x}$	$\vec{y}$	$\vec{z}$
User GPS	0	0	50			
RTK GPS	0	−31	50			
IMU	9.8	−25.5	0.5	$-\vec{x}_0$	$\vec{y}_0$	$-\vec{z}_0$
Velodyne	15	−8	31.6	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$
RGB	12	−44.5	46.5	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$
TIR	12	−44.5	40	$\vec{x}_0$	$\vec{y}_0$	$\vec{z}_0$

The locations of the sensor reference frames with respect to the vehicle's frame are illustrated in Figure 4 and defined in Tables 2 and 3 for the slightly different 2018 and 2019 sensor arrangements, respectively. The vehicle's reference frame $X_0Y_0Z_0$ has its origin at the platform surface, with the X_0 axis pointing in the motion direction. The GPS frames are placed at their respective antenna bases. Besides, the antenna's L1 band vertical offset is 5.43 cm. The origins for the Velodyne lidar and the cameras are at their corresponding optical centers. In the tables, frame orientations are expressed using the vehicle's reference frame unit vectors $\vec{x}_0$, $\vec{y}_0$ and $\vec{z}_0$. GPS frame orientations have been omitted because they are implicitly defined when transforming longitude-latitude coordinates, e.g., East-North-Up (ENU) or North-East-Down (NED).

3.3 *Data collection methodology*

We were authorized to conduct data acquisition runs during the full-scale SAR exercises on June 1, 2018 and on June 6, 2019. We recorded four sequences (two per exercise) with a total path length around 5.2 km and around 77 min of recording time. We drove the vehicle through different scenarios with the purpose of capturing participants, vehicles, and SAR activities and objects from a variety of distances and perspectives. Furthermore, we closed path loops so that data could be useful for SLAM research, and finished at the command post tent area. Additionally, we performed two post-exercise runs on July 25, 2018 and one pre-exercise run on June 5, 2019 in order to have comparable data of the empty site. These extra sequences add a path length around 4.9 km and 46 min.

Sensor data was captured by an on-board Intel NUC715BNKP, which included an i5-7260U processor

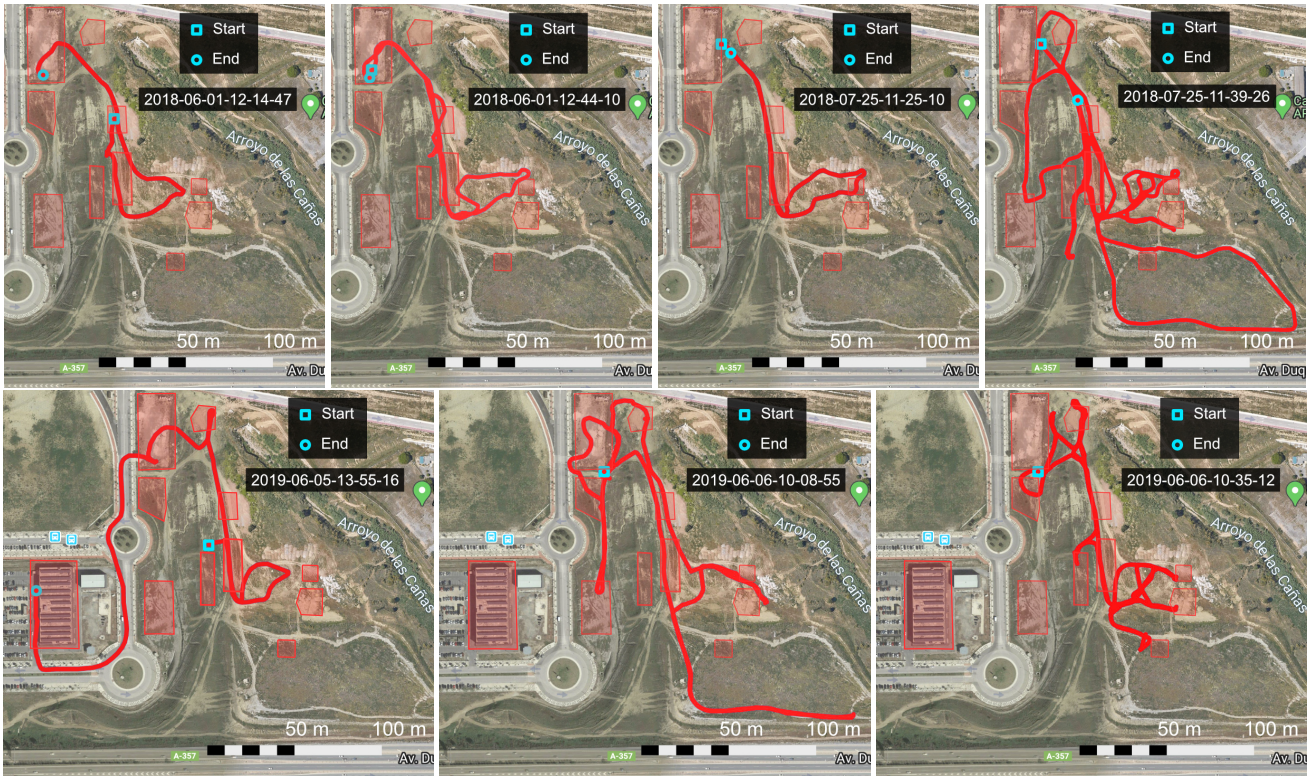

Figure 5. Run paths according to the standalone GPS. Top: 2018 datasets; bottom: 2019 datasets.

Table 4. Information for each dataset sequence.

	Sequence ID	Length (m)	Start time (UTC)	Duration (s)	Max / Min altitude (m)	Fix RTK msgs (%)	Avg. # of satellites	Horizontal speed Max/Avg. (km/h)	Temperature (°C)
SAR Exercise	2018-06-01-12-14-47	469	10:22:54	515.24	56.75 / 51.40	0.00	8.00	4.99 / 2.56	24.9
	2018-06-01-12-44-10	934	10:44:10	1201.35	57.02 / 50.91	0.00	8.00	5.76 / 2.79	24.9
	2019-06-06-10-08-55	1843	08:08:55	947.88	64.46 / 50.00	99.95	6.56	11.04 / 4.13	28.3
	2019-06-06-10-35-12	2000	08:35:12	1971.50	63.84 / 51.03	97.31	6.73	7.40 / 3.08	28.3
Empty site	2018-07-25-11-25-10	627	09:25:11	294.88	56.71 / 50.35	43.11	6.76	7.85 / 4.32	32.1
	2018-07-25-11-39-26	3320	09:39:27	2058.16	62.62 / 45.56	25.78	7.27	10.31 / 3.69	32.1
	2019-06-05-13-55-16	1029	11:55:16	434.23	57.93 / 45.61	99.58	6.16	12.89 / 5.48	27.4

with 8GB RAM and 256GB solid-state disk. We employed the Ubuntu 16.04 operating system running ROS Kinetic and recorded data with the rosbag ROS tool. In particular, the following ROS packages were used to access sensor data and to timestamp measurements based on system time:

- `axis_camera`, for the Axis P7214 server's MJPG streams (Gariepy 2018).
- `microstrain_3dmgx2_imu` for IMUs compatible with the microstrain 3DM-GX2 protocol (Leibs and Gassend 2013).
- `microstrain_mips`, for the 3DM-GX5-25 IMU (Vaughan 2020).
- `velodyne`, for Velodyne lidars (O'Quin 2019).
- `novatel_gps_driver_modified`, for publishing RMC and GGA NMEA messages of both JAVAD receivers. The original driver by Reed (2018) has been modified and it is available in the UMA-SAR Dataset website.

Additionally, the NTRIP client provided by RTKLIB (Takasu 2009) was employed to feed the RTK receiver with differential corrections. The `camera_calibration` ROS package was used to obtain the intrinsic parameter of the RGB camera (Bowman and Mihelich 2017). Images of an 11×7 chessboard (10 cm side) as well as calibration results are provided in the dataset website.

4 Dataset description

The complete dataset and accompanying software can be accessed through the website: www.uma.es/robotics-and-mechatronics/sar-datasets. This section describes the structure of the dataset, the data formats and the software tools for parsing and selecting data.

4.1 Dataset structure

The dataset has been divided into seven compressed files, each of which contains all raw data formats (i.e., human-readable text data and jpg images) for the corresponding run. Additionally, for each sequence we have provided the original rosbag files for the convenience of ROS-based software users. A map representation of the runs is shown in Figure 5. A preview of the RGB images as video stream format is also provided in the dataset website for each run. General information about the sequences is summarized in Table 4, including the sequence ID, the travelled distance, the

Table 5. Naming convention and human readable data format for the files in each sequence.

File name	Data	Format
image_{rgb/the}/ image_{rgb/the}-<timestamp>.jpg	RGB/Thermal images from the Oculus TI camera.	704×576 px JPG images with quality factor of 95 out of 100.
velodyne_points/ velodyne_points-<timestamp>.bin	Point cloud data from the Velodyne HDL-32. There is one 360° frame per file.	$N \times 4$ floating point matrix with N lidar points. Each column corresponds to $x/y/z$ (m) coordinates and reflectance (from 0 to 255) for each point.
imu/ data.txt	IMU data as text. There is one message per line.	Each line is structured as: timestamp (s), roll/pitch/yaw (rad), $x/y/z$ angular velocities (rad/s), $x/y/z$ linear accelerations ($\mathrm{m/s^2}$).
gps_{user/rtk}/ data.txt	Standalone/RTK GPS data with standard NMEA messages.	Each line contains the timestamps and GPGGA and GPRMC standard messages, providing positional information, ground velocity, satellite status.

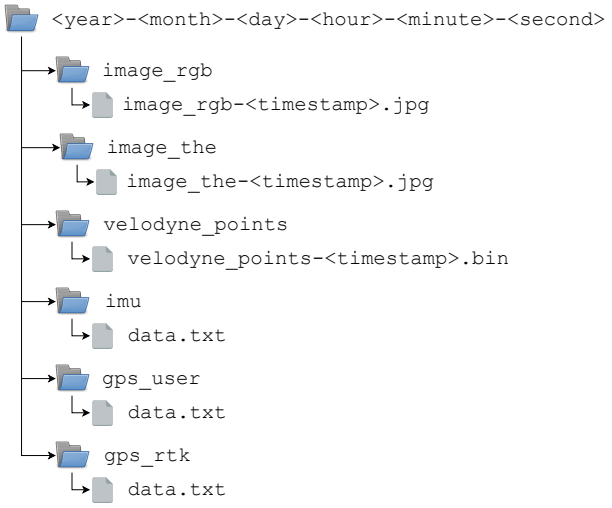

Figure 6. Data structure for each compressed file. Placeholders (<>) indicate variable data.

duration, maximum and minimum altitude, the percentage of fix RTK messages, the averaged number of satellites, the maximum and averaged horizontal speed considering only speeds over $0.2\,\mathrm{km/h}$, and the temperature records by the Spanish Meteorological Agency. The table also presents actual UTC starting time given by the GPS, which does not necessarily coincide with the PC system time used consistently in all dataset timestamps.

4.2 Raw data formats

The raw data structure aims to enhance human readability and indexing. Figure 6 shows the folder and file structure for a sequence. The root folder is tagged with the sequence ID, which indicates the recording date and a timestamp indicating starting time using the computer internal clock. Then, the data from each sensor is in a separate sub-folder. Image and lidar data are stored in multiple files with the following naming convention: `<device>-<timestamp>.<file extension>`. In contrast, IMU and GPS data are stored in single `data.txt` files, where each log is in a separate line starting with the timestamp and followed by the sensor reading.

The naming convention and a description for each sensor data is given in Table 5. The RGB and thermal images have been extracted in `jpg` format from the compressed images available in the rosbag files. Figure 7 shows image sample pairs from both cameras. As for the lidar frames, the timestamp in the file name indicates the moment of the first data point in the scan. The file stores the N valid range measurements of the frame packed into a binary file (`.bin`), which is common practice in lidar datasets (Maddern et al. 2017). Extraction of human readable data is straightforward with Python and Matlab code examples provided in a readme file. A sample point cloud from the Velodyne HDL-32 in 2019 is shown in Figure 8 along with the corresponding RGB and TIR images.

4.3 Rosbags

Apart from the raw data, we provide a rosbag for each sequence, which corresponds to the ROS standard messages obtained originally in the corresponding data collection run. The types of messages included in the .bag files are as follows:

- `sensor_msgs/CompressedImag`. RGB and TIR images from the Oculus-TI camera unit.
- `sensor_msgs/PointCloud2`. 3D point cloud from the Velodyne HDL-32.
- `sensor_msgs/Imu`. Raw IMU measurements from the 3DM-GX2 and 3DM-GX5-25 devices.
- `gps_common/GPSFix`. Additionally to longitude-latitude-altitude data, these messages include information of horizontal dilution of precision (HDOP), speed over the ground, track angle, time and number of satellites used.
- `novatel_gps_msgs/Gpgga`. NMEA GGA message from JAVAD User and RTK GPS recievers.
- `novatel_gps_msgs/Gprmc`. NMEA RMC message from JAVAD User and RTK GPS recievers.

4.4 Software tools

Two software tools for raw data sequence extraction and rosbag extraction, respectively, have been developed for the dataset. The tools allow adjusting the data to the preference and workflow of each user (e.g. allowing to reuse data loader scripts). The code with examples and detailed information about its usage is linked in the dataset website.

Figure 7. Four image pair samples from the RGB and TIR cameras. Left: 2018; right: 2019.

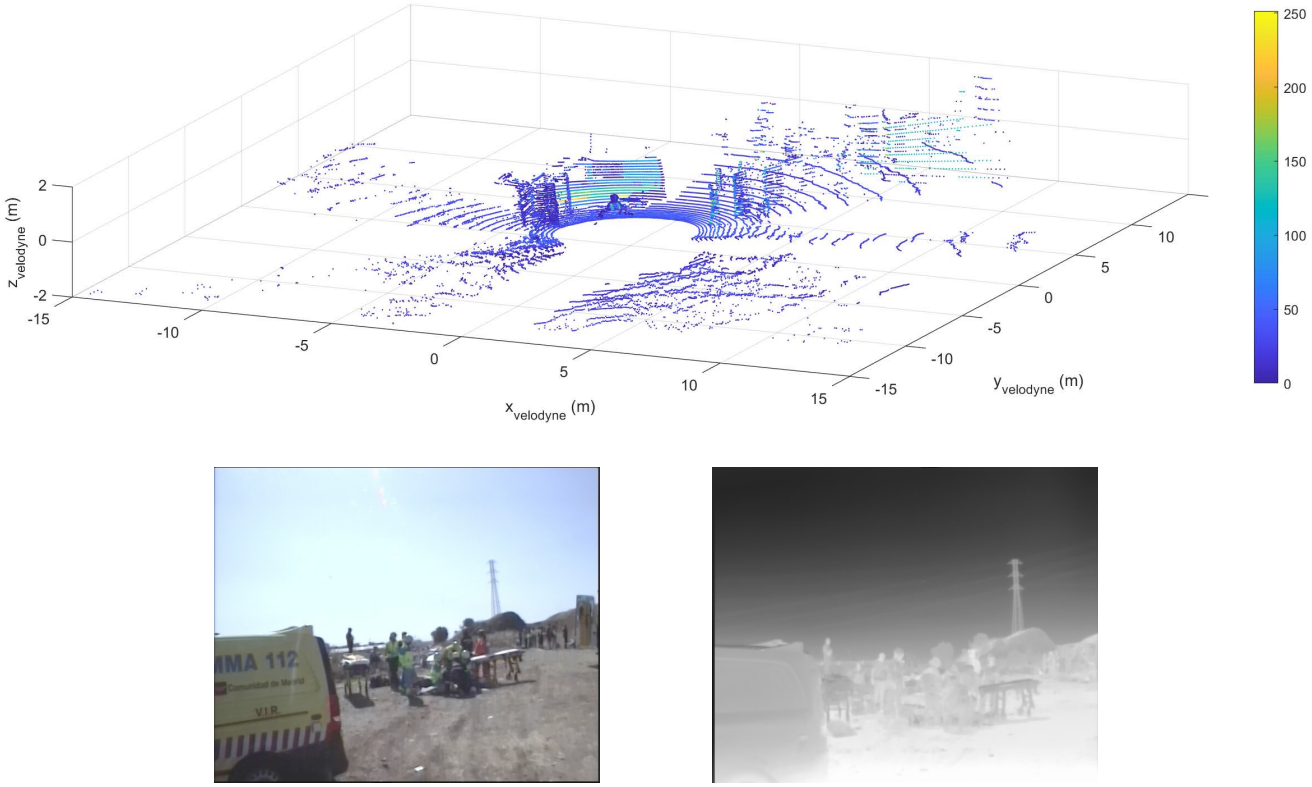

Figure 8. Multimodal data sample. Top: Point cloud from the Velodyne HDL-32 lidar. The color bar indicates reflectance. Bottom: Corresponding RGB and TIR images.

4.4.1 Sequence extraction. This tool has been developed for splitting the raw data sequences and selecting designated data sources. The software, developed using Python, uses a file specifying the sub-sequences to be extracted as the starting and the ending timestamps. This allows separating data for a specific purpose (e.g. only scenes with movement, or scenes with a particular object type, as people or cars), making it possible that the community of users can interchange or offer new splits using different criteria.

4.4.2 Rosbag extraction The rosbag extraction tool is a ROS package written in C++11 that includes the nodes used to extract the data from the rosbag files to raw data. The nodes were designed considering the easiness in their modification.

4.5 Disaster robotics data categories

The proposed dataset was captured from a manned vehicle emulating the path and sensor arrangement of a robotic UGV in fieldwork disaster exercises. This section relates the data offered in the UMA-SAR dataset with the data categories proposed by Murphy (2014) for data collection during fieldwork performance of disaster robots. In particular, six categories are defined:

1. Log of activity. The dataset website includes a text log file with the major events of each data capture

sequence with indications of time and traversed areas (as defined in Fig. 2). Besides, the GPS coordinates for each path are represented in Fig. 5, and corresponding operational and weather details are given in Table 4. Failures and bottlenecks are discussed in Section 5.

2. Context. Photographs in Fig. 1 provide context information. Section 3.1 describes the context of the fieldwork.

3. Robot's eye-view. A full video of the forward-looking RGB camera is provided for each sequence in the dataset website.

4. Robot's state. A detailed description of equipment, software, and collected data formats is given in Section 4, including two GPS receivers and IMU for vehicle localization.

5. External view. Some external video and photographs of the vehicle during the data capture missions are available in the dataset website.

6. Human-robot interaction. The vehicle was driven manually by a human operator.

5 Discussion

We believe that this dataset can be valuable for researchers working on different aspects of ground-based disaster robotics. Firstly, in order to solve a critical problem such as precise pose estimation, visual-inertial based odometry provides an accurate location with a very small drift. But SLAM still outperforms visual-inertial navigation in the study of sensor fusion in more challenging situations and the understanding of the environment topology thanks to loop-closure (or place recognition) (Cadena et al. 2016). In addition, SLAM can be useful for obtaining a global consistent map of the disaster site and ensuring a full coverage in an exploration task. In Liu et al. (2018) and Dubé et al. (2020) 3D lidar based SLAM methods on natural terrains and semi-unstructured environments are addressed, respectively, but not in collapsed, cluttered and unstructured outdoor environments where assumptions such as segments and planes are not valid. Recent works on large-scale and real-time visual SLAM show high quality maps and loop-closing in outdoor urban environments (Lynen et al. 2020) (Tanner et al. 2020), where GPS and 3D lidar are used as ground truth. Moreover, works in urban environments using vision, semantic mapping (Cadena et al. 2016) and including TIR images (Shin and Kim 2019) offer promising techniques for potential SLAM applications in disaster robotics. Nevertheless, semantic segmentation is a recent research area whose attention is currently focused on indoor (Milioto and Stachniss 2019) and urban environments (Zhang et al. 2018), but could be an interesting tool for disaster scenarios (Jeon et al. 2020). Throughout the UMA-SAR dataset, a wide range of representative SAR related object categories are present in overlapped RGB and TIR images. These images offer the opportunity to develop and train new machine learning models (Zhao et al. 2019) for scene understanding in SAR scenarios, where combining visible and thermal spectra allows the detection in hard visibility conditions (Bañuls et al. 2020). Among these representative categories, the localization of victims in a disaster site in life-critical SAR activities are highly important (Delmerico et al. 2019). This search of potential victims can be performed using multi-spectral images (Petříček et al. 2019) and including human body parts discovery (Oliveira et al. 2018). The aforementioned research topics are aimed at the disaster site recognition and mapping, where the location of first-response staff (and more information such as emergency vehicles, civilians, rubble mounds, tents and victims) is relevant for situational awareness, not only useful for robotic systems, but also for coordination personnel in charge of SAR missions. In fact, disaster robot applications that could benefit from novel intelligent solutions do not only include reconnaissance and mapping and search for survivors but also missions such as logistics, first medical assistance, casualty evacuation, and support for cooperative perception (Murphy 2014).

Nevertheless, the uniqueness of data collection in large-scale drills imposes error-prone conditions and severely limits the opportunities for solving technical problems on site (Leingartner et al. 2016). Thus, although useful, the dataset is affected by several issues. For instance, RGB images reflect vehicle vibrations and sudden turning motions, with occasional blurring that can be correlated with the IMU data. Moreover, the sensibility of the sensor suite to vibrations can also explain other imaging issues. Thus, in the 2019 TIR images, we detected band noise caused by wiring problems and high-amplitude vibrations. As for the RGB images, an internal mechanical deviation of the camera resulted on a small black portion of the casing's round window visible on the right corners of the June 1, 2018 images, as shown in Figure 7 (left). Besides, the same images show a printed text message in the right bottom corner. In addition, a general limitation of the RGB images is that we did not use a high definition camera. Initially, we were interested in using our standard definition Oculus device because it provides comparable RGB and TIR images, but also, we preferred not to release images where the identity of the responders (which included military units and other public security organizations) was clearly recognizable. A further limitation of the dataset has to do with RTK differential corrections. In 2018, corrections were not available due to interference affecting the UHF link to our base station, which could be explained by the concurrent use of the radio band by all participating organizations. As a result, both GPS receptors worked in standalone mode, so the dataset offers two redundant measurements with one-meter accuracy. In 2019, we coped with this problem by using a public positioning geodesic NTRIP server trough a 4G link, with only occasional losses of differential corrections that caused jumps in the information provided by the RTK GPS receiver. Deficient communications and GPS limitations are common problems faced by disaster robotics, so the multimodal measurements offered in the dataset may contribute to improve the robustness of localization and scene understanding in harsh SAR environments.

6 Summary

Our goal with this work was the public release of a multimodal dataset that can be useful for other researchers to test and improve their methods and algorithms. Thus, the robotics community can also profit from the exceptional opportunity to capture data from full-scale outdoor training exercises organized within our institution (UMA) for actual emergency response teams.

The dataset was recorded from a manned all-terrain vehicle equipped with a RGB camera, a thermal infrared camera (TIR) and a Velodyne HDL-32 3D lidar, as well as an inertial measurement unit (IMU) and two global positioning system (GPS) receivers as ground truth. The measurements capture unstructured terrain with rubble as well as a wide range of SAR-related object categories. Besides, overlapping TIR and RGB images can be suitable for machine learning and multi-spectral data fusion. Four data sequences, two from 2018 and two from 2019, were recorded during the exercises, with a total path length around $5.2\,\mathrm{km}$ and a total time around $77\,\mathrm{min}$. Additional sequences from the empty site provide an extra path length around $4.9\,\mathrm{km}$ and $46\,\mathrm{min}$.

The main aspects of the proposed dataset with respect to the review of previously published repositories from SAR ground-vehicles are following: i) It combines overlapping and synchronized RGB, thermal, and 3D lidar data modalities; ii) it has been conceived for cross-cutting robotics research areas including both SLAM and machine learning for scene understanding; and iii) it captures numerous instances of human subjects, objects and actions from the SAR domain.

The data is publicly available in human-readable raw format together with the original rosbag files. Besides, calibration data and software tools are also included. Since the UMA disaster exercises are conducted annually, our intention is to incorporate new datasets to our public repository in the future.

Acknowledgements

We are grateful to the Cátedra de Seguridad, Emergencias y Catástrofes, Universidad de Málaga, and in particular Prof. Jesús Miranda-Páez for providing access to the exercises and enabling the creation of this dataset. We would also like to thank all the organization staff and participants involved in the exercises. Finally, we wish to acknowledge the support of the UMA Robotics and Mechatronics Group.

Funding

This work has been performed in the frame of the project "TRUST-ROB: Towards Resilient UGV and UAV Manipulator Teams for Robotic Search and Rescue Tasks", funded by the Spanish Government (RTI2018-093421-B-I00) and project UMA18-FEDERJA-090 funded by the Andalusian Regional Government (Junta de Andalucía). David Morilla-Cabello was under a grant from the Spanish Government (Becas de colaboración 2019-2020).

References

Arnold S, Ohno K, Hamada R and Yamazaki K (2019) An image recognition system aimed at search activities using cyber search and rescue dogs. *Journal of Field Robotics* 36(4): 677–695. DOI:10.1002/rob.21848.

Balta H, Bedkowski J, Govindaraj S, Majek K, Musialik P, Serrano D, Alexis K, Siegwart R and De Cubber G (2017) Integrated data management for a fleet of search-and-rescue robots. *Journal of Field Robotics* 34(3): 539–582. DOI:10.1002/rob.21651.

Bañuls A, Mandow A, Vázquez-Martín R, Morales J and García-Cerezo A (2020) Object detection from thermal infrared and visible light cameras in search and rescue scenes. In: *IEEE International Symposium on Safety, Security, and Rescue Robotics*. pp. 380–386. DOI:10.1109/SSRR50563.2020.9292593.

Berrocoso M, Páez R, Jigena B and Caturla C (2006) The RAP Net: A geodetic positioning network for Andalusia (South Spain). *EUREF Publication* : 364–368.

Birk A, Schwertfeger S, Pathak K and Vaskevicius N (2009) 3D data collection at disaster city at the 2008 NIST response robot evaluation exercise (RREE). In: *IEEE International Workshop on Safety, Security and Rescue Robotics*. pp. 1–6. DOI:10.1109/SSRR.2009.5424164.

Bowman J and Mihelich P (2017) camera_calibration - ROS repository. URL http://wiki.ros.org/camera_calibration. Accessed July 29, 2020.

Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J, Reid I and Leonard JJ (2016) Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* 32(6): 1309–1332.

Choi Y, Kim N, Hwang S, Park K, Yoon JS, An K and Kweon IS (2018) KAIST multi-spectral day/night data set for autonomous and assisted driving. *IEEE Transactions on Intelligent Transportation Systems* 19(3): 934–948. DOI:10.1109/TITS.2018.2791533.

Dadwhal YS, Kumar S and Sardana HK (2020) Data-driven skin detection in cluttered search and rescue environments. *IEEE Sensors Journal* 20(7): 3697–3708. DOI:10.1109/JSEN.2019.2959787.

Delmerico J, Mintchev S, Giusti A, Gromov B, Melo K, Horvat T, Cadena C, Hutter M, Ijspeert A, Floreano D, Gambardella LM, Siegwart R and Scaramuzza D (2019) The current state and future outlook of rescue robotics. *Journal of Field Robotics* 36(7): 1171–1191. DOI:10.1002/rob.21887.

Droeschel D, Schwarz M and Behnke S (2017) Continuous mapping and localization for autonomous navigation in rough terrain using a 3d laser scanner. *Robotics and Autonomous Systems* 88: 104–115. DOI:10.1016/j.robot.2016.10.017.

Dubé R, Cramariuc A, Dugas D, Sommer H, Dymczyk M, Nieto J, Siegwart R and Cadena C (2020) SegMap: Segment-based mapping and localization using data-driven descriptors. *International Journal of Robotics Research* 39(2-3): 339–355. DOI:10.1177/0278364919863090.

Dubé R, Gawel A, Sommer H, Nieto J, Siegwart R and Cadena C (2017) An online multi-robot SLAM system for 3D lidars. In: *IEEE International Conference on Intelligent Robots and Systems*. pp. 1004–1011. DOI:10.1109/IROS.2017.8202268.

Duncan BA and Murphy RR (2014) Autonomous capabilities for small unmanned aerial systems conducting radiological response: Findings from a high-fidelity discovery experiment. *Journal of Field Robotics* 31(4): 522–536. DOI:https://doi.org/

10.1002/rob.21503.

Fernández-Lozano JJ, Mandow A, Martín-Guzman M, Martín-Avila J, Serón J, Martínez JL, Gomez-Ruiz JA, Socarrás-Bertiz C, Miranda-Paez J and García-Cerezo A (2018) Integration of a canine agent in a wireless sensor network for information gathering in search and rescue missions. In: *IEEE International Conference on Intelligent Robots and Systems*. pp. 5685–5690. DOI:10.1109/IROS.2018.8593849.

Gariepy R (2018) axis_camera - ROS wiki. URL https://wiki.ros.org/axis_camera. Accessed July 14, 2020.

Gawel A, Dubé R, Surmann H, Nieto J, Siegwart R and Cadena C (2017) 3D registration of aerial and ground robots for disaster response: An evaluation of features, descriptors, and transformation estimation. In: *IEEE International Symposium on Safety, Security and Rescue Robotics, Conference*. pp. 27–34. DOI:10.1109/SSRR.2017.8088136.

Google (2020) Aerial view of the UMA search and rescue experimental area. URL https://goo.gl/maps/EC2v2y1LtbRvBu4M7. Accessed August 04, 2020.

HRAIL (2020) HRAIL repository. Humanitarian robotics and AI laboratory at Texas A&M university. URL http://hrail.crasar.org. Accessed July 29, 2020.

Jeon HG, Im S, Lee BU, Choi DG, Hebert M and Kweon IS (2019) DISC: large-scale virtual dataset for simulating disaster scenarios. In: *IEEE International Conference on Intelligent Robots and Systems*. pp. 187–194. DOI:10.1109/IROS40897.2019.8967839.

Jeon HG, Im S, Oh J and Hebert M (2020) Learning shape-based representation for visual localization in extremely changing conditions. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 7135–7141. DOI:10.1109/ICRA40945.2020.9196842.

Krišto M and Ivăsić-Kos M (2019) Thermal imaging dataset for person detection. In: *International Convention on Information and Communication Technology, Electronics and Microelectronics*. pp. 1126–1131. DOI:10.23919/MIPRO.2019.8757208.

Kubelka V, Reinstein M and Svoboda T (2019) Tracked robot odometry for obstacle traversal in sensory deprived environment. *IEEE/ASME Transactions on Mechatronics* 24(6): 2745–2755. DOI:10.1109/TMECH.2019.2945031.

Leibs J and Gassend B (2013) microstrain_3dgx2_imu - ROS wiki. URL http://wiki.ros.org/microstrain_3dmgx2_imu. Accessed July 14, 2020.

Leingartner M, Maurer J, Ferrein A and Steinbauer G (2016) Evaluation of sensors and mapping approaches for disasters in tunnels. *Journal of Field Robotics* 33(8): 1037–1057. DOI:10.1002/rob.21611.

Liu Z, Chen H, Di H, Tao Y, Gong J, Xiong G and Qi J (2018) Real-time 6D lidar SLAM in large scale natural terrains for ugv. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. pp. 662–667.

Lorenz P and Steinbauer G (2018) The RoboCup Rescue victim dataset. In: *IEEE International Symposium on Safety, Security, and Rescue Robotics*. pp. 1–6. DOI:10.1109/SSRR.2018.8468605.

Lynen S, Zeisl B, Aiger D, Bosse M, Hesch J, Pollefeys M, Siegwart R and Sattler T (2020) Large-scale, real-time visual–inertial localization revisited. *The International Journal of Robotics Research* 39(9): 1061–1084. DOI:10.1177/0278364920931151.

Maddern W, Pascoe G, Linegar C and Newman P (2017) 1 year, 1000 km: The Oxford RobotCar dataset. *International Journal of Robotics Research* 36(1): 3–15. DOI:10.1177/0278364916679498.

Milioto A and Stachniss C (2019) Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using CNNs. In: *2019 International Conference on Robotics and Automation (ICRA)*. pp. 7094–7100.

Mishra B, Garg D, Narang P and Mishra V (2020) Drone-surveillance for search and rescue in natural disaster. *Computer Communications* 156: 1–10. DOI:10.1016/j.comcom.2020.03.012.

Murphy RR (2014) *Disaster Robotics*. The MIT Press.

Murphy RR, Duncan BA, Collins T, Kendrick J, Lohman P, Palmer T and Sanborn F (2016) Use of a small unmanned aerial system for the SR-530 mudslide incident near Oso, Washington. *Journal of Field Robotics* 33(4): 476–488. DOI:10.1002/rob.21586.

Ohno K, Tadokoro S, Nagatani K, Koyanagi E and Yoshida T (2010) Trials of 3-D map construction using the tele-operated tracked vehicle Kenaf at Disaster City. In: *IEEE International Conference on Robotics and Automation*. pp. 2864–2870. DOI:10.1109/ROBOT.2010.5509722.

Oliveira GL, Bollen C, Burgard W and Brox T (2018) Efficient and robust deep networks for semantic segmentation. *The International Journal of Robotics Research* 37(4-5): 472–491. DOI:10.1177/0278364917710542.

O'Quin J (2019) Velodyne - ROS wiki. URL http://wiki.ros.org/velodyne. Accessed July 14, 2020.

Pathak K, Birk A, Vaskevicius N, Pfingsthorn M, Schwertfeger S and Poppinga J (2010) Online three-dimensional SLAM by registration of large planar surface segments and closed-form pose-graph relaxation. *Journal of Field Robotics* 27(1): 52–84. DOI:https://doi.org/10.1002/rob.20322.

Pellenz J, Lang D, Neuhaus F and Paulus D (2010) Real-time 3D mapping of rough terrain: A field report from Disaster City. In: *IEEE Safety Security and Rescue Robotics*. pp. 1–6. DOI:10.1109/SSRR.2010.5981567.

Petříček T, Šalanský V, Zimmermann K and Svoboda T (2019) Simultaneous exploration and segmentation for search and rescue. *Journal of Field Robotics* 36(4): 696–709. DOI:10.1002/rob.21847.

Reed JP (2018) novatel_gps_driver - ROS repository. URL http://wiki.ros.org/novatel_gps_driver. Accessed July 14, 2020.

Shin YS and Kim A (2019) Sparse depth enhanced direct thermal-infrared slam beyond the visible spectrum. *IEEE Robotics and Automation Letters* 4(3): 2918–2925. DOI:10.1109/LRA.2019.2923381.

Shrewsbury B, Henkel Z, Kim CY and Murphy RR (2013) RESPOND-R test instrument: A Summer institute 2013 case study. In: *IEEE International Symposium on Safety, Security, and Rescue Robotics*. pp. 1–6. DOI:10.1109/SSRR.2013.6719369.

Svoboda T (2017) ROS bag-files. TRADR: long term human-robot teaming for robot assisted disaster response. URL http://www.tradr-project.eu/resources/datasets/. Accessed July 7, 2020.

Takasu T (2009) RTKLIB: Open source program package for RTK-GPS. In: *Free and Open Source Software for Geospatial*

(FOSS4G). pp. 852–861.

Tanner M, Piniés P, Paz LM, Ştefan Săftescu, Bewley A, Jonasson E and Newman P (2020) Large-scale outdoor scene reconstruction and correction with vision. *The International Journal of Robotics Research* DOI:10.1177/0278364920937052.

Vaughan J (2020) fiducials - GitHub repository. URL https://github.com/ros-drivers/microstrain_mips. Accessed July 14, 2020.

Xu Z, Zhuang J, Liu Q, Zhou J and Peng S (2019) Benchmarking a large-scale FIR dataset for on-road pedestrian detection. *Infrared Physics and Technology* 96: 199–208. DOI:10.1016/j.infrared.2018.11.007.

Zhang X, Chen Z, Lu D and Li X (2018) Real-time semantic segmentation for road scene. In: *International Conference on Advanced Robotics and Mechatronics*. pp. 19–23.

Zhao Z, Zheng P, Xu S and Wu X (2019) Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems* Early access: 1–21. DOI:10.1109/TNNLS.2018.2876865.