Computer science
Case study: Rescue robots

For use in May and November 2024

Instructions to candidates

- Case study booklet required for higher level paper 3.

Scenario

BotPro is a company that makes robots for various industrial applications. However, in the past decade the company has focused on making rescue robots. A rescue robot is designed to help with the search and rescue of humans after a disaster, such as an earthquake or tsunami. These robots
5   may assist the efforts of rescue teams by searching and mapping areas, assessing damage, removing debris, delivering supplies, and evacuating casualties.

Last year, *BotPro* had to send one of their rescue robots into a large factory after a fire. On a second occasion, the rescue robot was sent to a factory that was damaged by an earthquake. On both occasions, *BotPro* found that the rescue robot performed below expectations. The robot
10   had trouble navigating the factories and reaching the exact location of survivors.

The chief executive officer (CEO) of *BotPro* started to investigate the reasons for this poor performance. He knew that the same robot had been deployed to assist in the rescue of people from a stadium after structural damage and that the operation was successful. So what was different now?

15   After investigation, the main difference was found to be in the accuracy and strength of the *global positioning system (GPS) signal* inside the factories. The strength of the satellite signal received was insufficient, resulting in errors in calculating rescue robot's exact location. The factories were also an unknown environment to the rescue robot, and the damage caused meant that the layout of the factories was different to any floor plans that existed.

20   # Problems to be addressed

The CEO of *BotPro* commissioned work on redesigning a rescue robot with the capability to explore, map, and survey the interior of unknown and damaged buildings. It was important to the CEO that this new rescue robot was cost effective and that the focus of the new design was on building efficient algorithms to improve robot performance and not on the mechanical components
25   of the robot.

The design team made a list of problems that needed to be addressed:
1.   **Accurate mapping of the area.** The rescue robot needs to find its way inside buildings and reliably operate in a *GPS-degraded environment* or *GPS-denied environment* and in the absence of an exact map.
30   2.   **Navigation in a dynamic and unknown environment.** The rescue robot has to navigate in an unknown environment where structures may have been damaged and changed due to the disaster.
3.   **Finding survivors.** The rescue robot is required to detect debris and humans under different light conditions, including darkness, deal with occlusion by other objects, and recognize
35   deformed shapes.
4.   **Communication.** The rescue robot needs to communicate with the rescue team outside the damaged space. The rescue robot may need to rely on large databases and the processing power of central computers.

Technologies

40 The rescue robot must be able to move around the environment. To move around the environment successfully, a rescue robot needs maps. Most commonly, GPS signals are used for navigation in open spaces, as the signals received are accurate. This is not the case for indoor spaces.

The design team at *BotPro* therefore needs to create a visual situational awareness and are looking at *computer vision* techniques to allow the rescue robot to produce a map of its
45 surroundings and recognize objects not seen before. At the most basic level, the rescue robot needs to be equipped with an *odometry sensor* and a single camera.

Computer vision

Computer vision refers to technologies that allow a device to "see", that is, sense the environment around it, including both static and dynamic objects.

50 Two of the subdomains of computer vision technologies that the design team at *BotPro* are exploring for the new rescue robot include *visual simultaneous localization and mapping (vSLAM)* and pose estimation.

Visual simultaneous localization and mapping (vSLAM)

Simultaneous localization and mapping (SLAM) is used for estimating sensor motion and
55 reconstructing the structure of an unknown environment. SLAM can be performed using different kinds of sensors, such as cameras for visual data collection and radar, sonar, and *light detection and ranging (LIDAR)* for non-visible data, with basic positional data collected using an *inertial measurement unit (IMU)*. An IMU is a combination of sensors, such as an accelerometer, gyroscope, and magnetometer, that capture the data about a device's location
60 in a three-dimensional (3D) space. If installed on the rescue robot, it would measure the translational and rotational movements of the robot.

SLAM techniques that use a camera as the basic visual input are known as vSLAM or visual SLAM.

vSLAM algorithms will allow the rescue robot to dynamically build a map while keeping track of
65 its own location and orientation as it moves around an unknown environment. The objective is to combine the *dead reckoning data* provided by the odometry sensor and images from the camera as an input to undertake localization and map building.

vSLAM works in indoor spaces and provides greater accuracy than GPS, giving it an advantage over GPS. **Figure 1** shows the vSLAM process chart.

Figure 1: vSLAM process chart

| Input data from camera / other sensors | → | Feature detection and extraction | → | Data association | → | Update pose and map |

Turn over
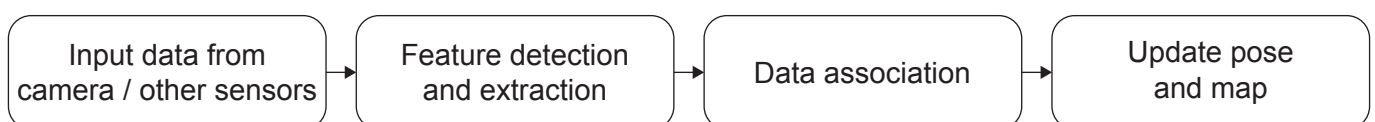
70   The vSLAM algorithm has a number of modules, the three main ones being *tracking*, *local mapping*, and *loop closure*. They are shown in **Figure 2**.

Figure 2: The three main modules of the vSLAM algorithm
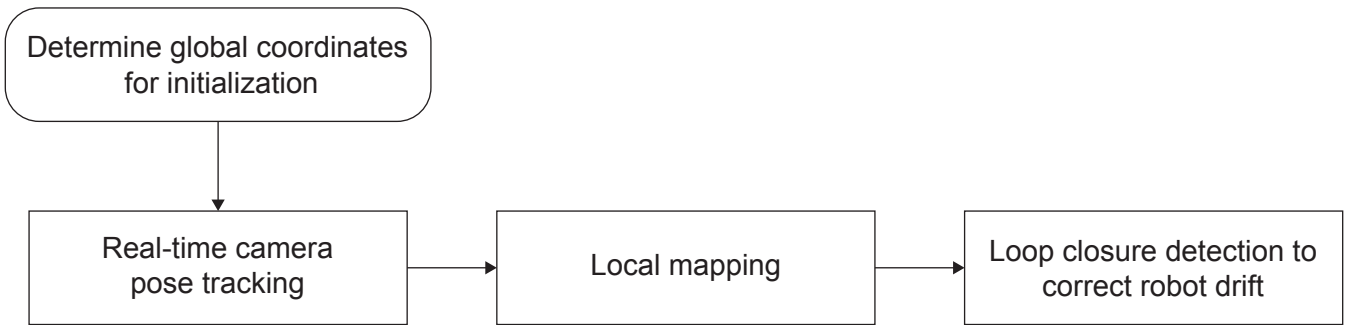
The rescue robot uses the three main modules of the vSLAM algorithm to initialize its position according to the available data, track its progress while analysing new sensor data, and create a map of the environment.

75   As the measurement of features, such as doors, wall corners, and the location/pose of the people in the damaged structure, increases over time, the environment representation needs to be optimized, which requires significant computational processing capabilities (see **Figure 3**). Thus, a balance is needed between the efficiency of the optimizing algorithms and the accuracy of the map that is created. The two techniques used for *optimization* are *bundle adjustment* and
80   *keyframe selection*.

Figure 3: Flowchart of a typical vSLAM process
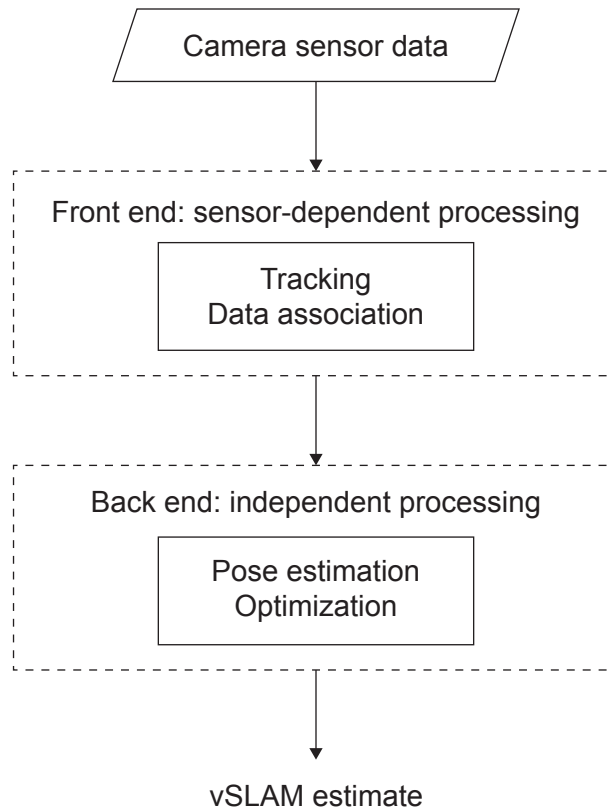
When tracking fails, vSLAM executes a module for *relocalization*, as the rescue robot needs to relocalize itself. As the rescue robot moves through a space, it may drift and lose itself, however, at this time *global map optimization*, which uses a loop closure technique, is executed.

85   The design team at *BotPro* realized that vSLAM is a low-cost and robust algorithm that can handle dynamic changes in the environment. However, it is important to know the performance, advantages, and limitations of the vSLAM algorithm.

The design team are also exploring the idea of using multiple cameras in the new rescue robot. If they implement this idea, a different image acquisition scheme may be used to optimize the use of visual information.

90   ## Pose estimation

BotPro's new rescue robot needs to know the pose of a person in order to perform correct rescue operations and therefore needs to be able to estimate the configuration of human body parts.

Pose estimation refers to the computer vision technique of estimating the position and orientation of an object or a human relative to the camera in a real-world space. This is usually done by
95   identifying, locating, and tracking a number of *key points* on an object or a human in an interactive environment. Key points are spatial locations or points in an image that define what stands out in the image captured by the camera.

Inanimate objects are rigid, so their key points are typically corners, edges, or other significant features. Models used to identify key points on inanimate objects are called *rigid pose estimation*
100   *(RPE)* models.

Humans are flexible, so their key points are the head and major joints such as the neck, shoulders, elbows, wrists, and knees. Models used to identify key points on humans are called *human pose estimation (HPE)* models (see **Figure 4**).

Figure 4: A kinematic representation of a human body
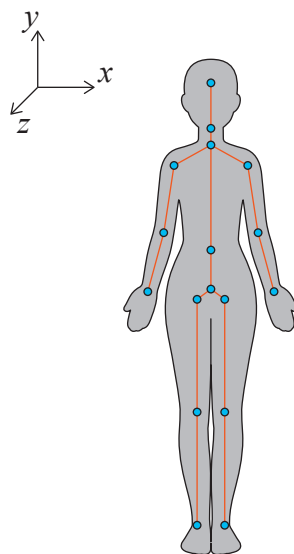
Key points can be used to describe the pose of a human in two dimensions. A valid connection
105 between points is known as a pair. This two-dimensional (2D) image of key points and pairs is
then transformed into a 3D pose model that enables the prediction of the accurate positioning
of a person (see **Figure 5**). HPE does not recognize who is in the image; it only estimates the
body parts.

Figure 5: 3D human pose estimation
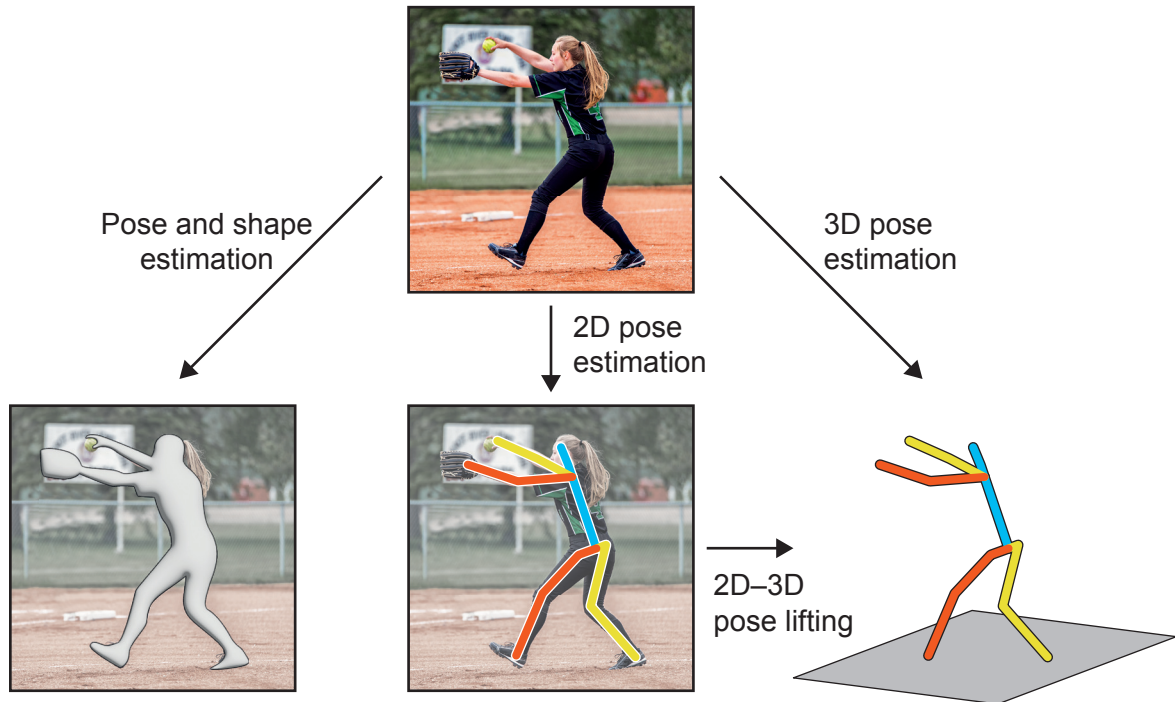
HPE methods are classified into two types: bottom-up methods and top-down methods.

110 HPE methods also differ depending on the number of people being tracked. Single-person
pose estimation is easier than multiple-person pose estimation.

The question facing the design team at *BotPro* is, does the rescue robot need a highly accurate
image of the people it is trying to find and rescue? Achieving higher accuracy would imply an
increase in computational complexity, which would lead to an increase in inference delay in
115 real-time rescue operations.

Pose-tracking algorithms face several challenges, including background scene handling,
self-occlusion of body parts, tracking in diverse light conditions, and multiple-*object occlusion*.

Rescue robots must undertake pose estimation in real time, with or without internet connectivity,
so the design team at *BotPro* might consider using lightweight edge devices on the new rescue
120 robot that implement *edge computing*. The challenge then would be to enable the rescue robots
to work with low-power devices.

The new rescue robot will also need two-way communication with the human rescue team.
The design team at *BotPro* are aware that access to data/images sent by the rescue robot
should be limited to authorized persons to maintain the security and privacy of data.

125 Social and ethical issues

The use of rescue robots has several advantages. For example, rescue robots can be sent to places that are too dangerous for human rescue workers. A rescue robot is designed to assist humans, saving human rescue workers from exhaustion and trauma. Furthermore, rescue robots can be used to gather feedback from mock deployments that can be used to train human rescue
130 workers about safety and efficiency in disaster management teams.

However, rescue robots require a large amount of resources to build and maintain, yet they may have limited capability. As such, only more economically developed countries can afford to spend time and money on building and using them.

Some questions about the social and ethical aspects of using rescue robots still remain:
135 • Should rescue robots be used to perform tasks that would be too dangerous for humans in a disaster or similarly unstable environment?
• Are robots safe? Who is accountable if a rescue robot causes harm to the person in need of help due to a miscalculation of the person's pose?
• Will survivors be comfortable being helped by a robot and not another compassionate
140 human being?

Further research

Future advances by *BotPro* would explore *sensor fusion models* to provide greater confidence and accuracy.

The company would also like to make multiple rescue robots that can work collaboratively. This will
145 require them to identify the algorithms that can be used to enable collaboration not just between the robots but also with their human operators.

Technologies like intelligent network infrastructure that can be dynamically enhanced and extended by edge nodes could serve as the backbone. The ability to be connected in any place and at any time would allow a necessary level of autonomy and collective exchange of information
150 among robots, humans, infrastructure and related applications. However, this could require more advanced edge computing and distributed machine learning.

Building a rescuer graphical user interface (GUI) that could map a disaster zone in real time to prepare the human rescue team if they wanted to enter the disaster zone is another area that the design team at *BotPro* have identified to improve their rescue efforts.

155 Challenges faced

In the development of the new rescue robot, the design team at *BotPro* face a number of challenges:
• Understanding how vSLAM navigates an environment with unknown obstacles and contours.
• Minimizing the time rescue robots spend scanning and learning an environment.
160 • Estimating the pose of people despite varying light and environmental conditions and body-part or multiple-object occlusion.
• Updating existing maps in a dynamically changing environment, such as an earthquake where rubble is still shifting.
• Developing an understanding of the ethical considerations of using autonomous robots in
165 life-and-death situations.

Turn over

Additional terminology

Bundle adjustment
Computer vision
Dead reckoning data
Edge computing
Global map optimization
Global positioning system (GPS) signal
GPS-degraded environment
GPS-denied environment
Human pose estimation (HPE)
Inertial measurement unit (IMU)
Keyframe selection
Key points/pairs
Light detection and ranging (LIDAR)
Object occlusion
Odometry sensor
Optimization
Relocalization
Rigid pose estimation (RPE)
Robot drift
Simultaneous localization and mapping (SLAM)
Sensor fusion model
Visual simultaneous localization and mapping (vSLAM) modules:
    Initialization
    Local mapping
    Loop closure
    Relocalization
    Tracking

Some companies, products, or individuals named in this case study are fictitious and any similarities with actual entities are purely coincidental.

References:

Figure 4   Xiaopeng JI, Qi FANG, Junting DONG, Qing SHUAI, Wen JIANG, Xiaowei ZHOU. A survey on monocular 3D human pose estimation. *Virtual Reality & Intelligent Hardware*, 2020, 2(6): 471—500 DOI: 10.1016/j.vrih.2020.04.005. Image adapted.

    Sudowoodo, n.d. [Female body chart] [online]. Available at: https://www.gettyimages.co.uk/detail/illustration/male-and-female-body-chart-royalty-free-illustration/1160619280?phrase=body&adppopup=true [Accessed 18 October 2022]. Source adapted.

Figure 5   Xiaopeng JI, Qi FANG, Junting DONG, Qing SHUAI, Wen JIANG, Xiaowei ZHOU. A survey on monocular 3D human pose estimation. *Virtual Reality & Intelligent Hardware*, 2020, 2(6): 471—500 DOI: 10.1016/j.vrih.2020.04.005. Image adapted.

    Kat72, n.d. [Teenage girl pitching fastball] [online]. Available at: https://www.gettyimages.co.uk/detail/photo/teenage-girl-pitching-fastball-royalty-free-image/700576396 [Accessed 18 October 2022]. Source adapted.