https://doi.org/10.2352/ISSN.2470-1173.2019.7.IRIACV-463

Hazmat label recognition and localization for rescue robots in disaster scenarios

Raimund Edlinger, Gerald Zauner, Michael Zauner; University of Applied Sciences Upper Austria, Wels, Austria

Abstract

Firefighting and rescue live victims operations are inherently dangerous, but the imminent danger of release of a hazardous substance creates an additional risk. Thus, identification of hazardous materials during robot assisted search and rescue missions can help e.g. firefighters or rescue teams to improve such rescue operations. The paper deals with the development of such a robotic machine vision system for hazmat label recognition. Classical computer vision methods but also state-of-the-art deep learning based detection algorithms were implemented and evaluated. Special focus was put on the robustness of detection and recognition with limited hardware resources and the influence of background image structures and light conditions.

Introduction

The automation of tasks is no longer only driven by industry. More and more everyday tasks are carried out by partly static and partly mobile machines. The aim is to have monotonous or repetitive tasks carried out by robots. This usually increases not only the throughput but also the quality of the work.

In recent years, efforts have also been made to support rescue forces in their difficult and dangerous tasks. In order to advance research in this area, robot competitions have been launched to target precisely such scenarios. These efforts were reinforced by the incident at the Fukushima nuclear power plant in 2011. Specifically, the Rescue League was established at the end of the 1990s as part of the RoboCup challenge. The idea for this league came after the earthquake in Kobe City [1]. Subsequently, the test methods for such competitions became more and more standardized in order to make the robots comparable with each other [2]. In recent years, also the DARPA Robotics Challenge has provided sensational results in the field of robot development [3].

In these challenges various test methods are provided that measure robot manoeuvring, mobility, sensors, energy, radio communication, dexterity, durability, logistics, semi- and full autonomy functions and operator proficiency. These tests are a possibility to evaluate also quantitatively, whether robotic systems are capable and reliable enough to perform operational tasks. This is very important for first responders, as they should actively use these machines. Another benefit, from a technical point of view, is to push integration of on board sensors and intelligent controls to improve remote operator capabilities.

The World Robot Summit (WRS) in Tokyo also belongs to the group of robot competitions described above. Here, among other things, special operations are being tested which stand in connection with accidents in nuclear power plants. The WRS sees itself as a worldwide platform to share knowledge for robot developers, researchers, official government agencies and private industry [4]. A lively exchange is to take place between these groups and thus the development of modern robot technology is to be advanced. In addition to the actual competition, there are also exhibitions related to robotics.

At the WRS, robot developers have to solve various vision and dexterity tasks. These are tested in different categories and according challenges:

1. *Industrial Robotics:* this challenge is addressed to robots in manufacturing and component construction.
2. *Service Robotics:* here the tasks in the area of cleaning habitats, housekeeping and interaction between robots and humans are set.
3. *Disaster Robotics:* in this category, robots must perform preventive tasks in a factory environment to prevent an accident. In addition, the robots are intended to support emergency personnel in scenarios like tunnel accidents, after earthquakes or incidents in power plants.

The main objective of robotic competitions is to hold challenging and fair competitions that inform teams of the tasks necessary to be effective for the task forces and first responders. The challenge progress of each robot system must also be measured to highlight breakthrough skills that respondents can understand and appreciate. Ten or more successful repetitions typically indicate a reliable skill.

In this context, special focus is on treatment of hazardous materials, which are usually stored in a variety of containers of different sizes, shapes and types. Typically, they are transported and stored in these trailers but also in the form of sacks, cartons and barrels up to tanks and bottles. When looking at the recognition and identification process, first responders should be alert when arriving at an incident, i.e. it should be the responsibility of robots to identify such hazardous substances as quickly and accurately as possible to inform first responders of possible dangers. The robot system will therefore serve as a digital helper with built-in autonomous systems such as robust recognition and identification of hazard materials to ensure safe operation and recovery in different environments.

One common labelling system, which is used to identify hazardous materials, is the NFPA 704 system. It groups all chemical hazards (health, flammability, reactivity) and special hazards into a single type of sign. Figure 1 shows an example of an NFPA rating criterion and a typical NFPA 704 label, where different types of hazards are represented by different colors. A numerical rating system is provided to evaluate various hazards, where 0 represents the lowest hazard and 4 the most dangerous.

Thus, the purpose of the work done is to develop and evaluate a mobile robot based system for hazardous materials detection and localization. The objective is to develop a robust vision system which robustly identifies the nine hazard classes and the hazards associated with each hazard class. The location and quantity of hazardous substances is also important information for the investigation and recovery of such materials

Figure 1: Example of a NFPA 704 Label [18]

Mobile Robot System: Hardware and Computing

In order to solve all these mobility, dexterity and exploration tasks, teams need a high mobility robot equipped with powerful onboard computing for autonomous operation.

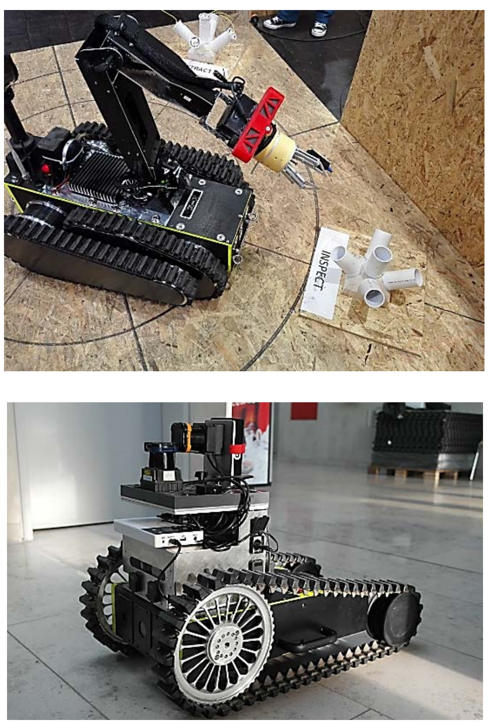

Figure 2: UGVs for rescue and exploration applications.

The Unmanned Ground Vehicles (UGVs) used in this work are self-developed rescue robots which are equipped with a chain drive (Figure 2). The robots are connected to an operator station via 5 GHz WLAN and via 433 MHz. They are primarily developed for the use in the field of security and emergency applications. The aim is to build autonomous and tele-operated robots, which are able to drive through an unstructured environment and search for potential human victims. This includes generating a map of the explored environment and characterizing and locating victims as well as 'recognizing' dangerous situations and hazardous materials (e.g. caused by fire and gas).

The computing platforms used in our tests are a NVIDIA Jetson TX2 mobile computer (onboard the robot) and a laptop with a 2.6 GHz i7-6600U CPU (remote operator station). To create a map of the current scenario, different 2D- and 3D-sensors are mounted on the robot. A realsense™ camera system in front of the robot arm is used for readiness tests in identification- and dexterity tasks. A Velodyne VLP-16 Lidar is used for 2D and 3D mapping of the environment and to localize/visualize the hazmat labels in a 3D map (Figure 3).

Figure 3: Typical identification tasks

Related Work

There are several computer vision algorithms, which are used to perform object detection in images. Some examples for these methods are image segmentation with blob analysis, simple pattern matching, or feature extraction based algorithms like SIFT, SURF, GLOH, HOG, BRIEF, etc. supported by classification algorithms like SVMs (support vector machines) [5-16].

New deep learning approaches such as Convolutional Neural Networks (CNNs) have now mostly replaced traditional methods and provide a new range of computer vision tools. It was the emerging technology of GPUs that finally enabled deep learning to become one of the most powerful computer vision tools for object detection today.

In this work, we first applied the classic detector algorithm from Viola and Jones, which was further improved by Lienhart [14], and additionally two keypoint-feature based algorithms, namely Speeded Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) [9,10,13]. Finally, we implemented a new and fast CNN based image object detector named YOLOv3 [17].

After the successful detection of a hazmat label, the detection location must subsequently be marked on a generated map. So-called SLAM (synchronous localization and mapping) algorithms are available for this purpose, which usually create such maps based on 3D sensors. Here we extrinsically calibrate the RGB camera with a 3D depth camera (realsense™ camera) to obtain the relative distance of the detected hazmat label to the robot. Using the SLAM algorithm for self-localization of the robot, this relative position can then be related to the map created.

Hazmat Label Detection

In order to solve all the rescue tasks described above, not only a high mobility of the robots but also a very robust perception of the environment is required. Camera systems are increasingly being used as a cost-effective supplement to 3D laser sensors. Our mobile robot uses a low-cost RGB-camera (640x480 pixels) and ROS-based (Robot Operating System [20]) implementations of the above mentioned keypoint-detectors: *AffineSIFT* [9] and *SURF* [10]. We also used a Qt/OpenCV based software named *Find Object* by Mathieu Labbé (http://introlab.github.io/find-object/). We then compared these results with those of a recently developed algorithm based on deep learning *(YOLOv3)* [17].

Another focus, beside robust recognition results, is on the necessary computing capacity needed for these different algorithms (due to the correspondingly compact design and the limited onboard power supply). The aim is to evaluate qualitatively the differences between keypoint-feature based detection algorithms and newly developed deep learning algorithms, also with respect to possible video real-time performance. Especially, poor or changing lighting conditions and slight changes of perspective should not affect the detection results. In addition, the vision system should cope with large variations in object distances (large scale invariance).

Method I: "Classic" Vision Approaches

SURF is a keypoint detector that is invariant to image scaling and rotation and builds on the widely used SIFT detector, but is considered to be much faster. Using integral images, the SURF algorithm applies average filters instead of the Gaussian filters of the SIFT detector. This speed-up plays an important role especially in the video real-time application considered.

Figure 4: Typical result of the SURF detector (OpenSurf) - only a very little number of keypoints can be detected.

As exemplarily shown in figure 4, the OpenSURF detector hardly recognizes any points under difficult light conditions, whereas the AffineSIFT algorithm assigns nearly all feature points correctly (Figure 5, left) – even when the camera perspective is changed (Figure 5, right).

Although the real-time performance of these algorithms was satisfactory, we soon encountered severe problems with the robustness of detection. In real world test drives in our test arena these approaches could simply not achieves acceptable results.

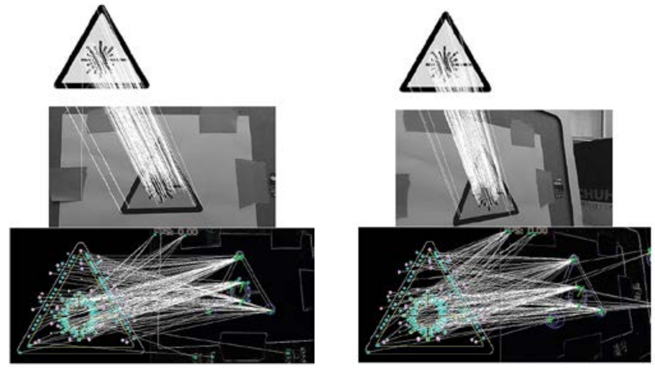

Figure 5: Comparison of AffineSIFT: front view (left) and object rotated (right).

The decision then was to implement deep learning methods, as they have proven to be extremely robust in many other real world applications, such as autonomous driving. In addition, we argued that the vision system should be very flexible so that it can be adapted easily to new scenarios by a simple learning process (e.g. for different robot challenges).

Method II: Deep Learning

A general challenge in deep learning is the enormous amount of learning data needed to produce good results. CNNs are supervised learning approaches, i.e. labeled images that constitute the basic truth must be initially provided to train the neural network.

Training Data and Image Labelling

We used a software named *darklabel (*https://darkpgmr.tistory.com/16*)* for image labelling. This software provides an easy to use GUI. It supports basic object tracking functionalities which substantially simplifies the otherwise time consuming work of object labelling in images as a once marked image object can be recognized again in the following images and a new drawing of the object borders is done automatically.

We defined 12 hazmat object classes (Table 1) and we used about 5500 training images (450 images/class) which we made from videos of real hazmat labels placed on a wall. In this context, we have also consciously taken care to include appropriately rotated and blurred images in the training data set.

Table 1: Hazmat Label Classes

- corrosive_8
- dangerous_4
- explosive_S_1_4_1
- flammable_liquid_3
- flammable_solid_4
- infectious_substance_6
- inhalation_hazard_2
- non-flammable_gas_2
- organic_peroxide_5_2
- oxidizer_5_1
- radioactive_II_7
- spontaneously_combustible_4

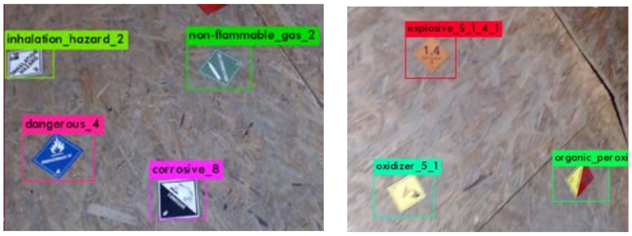

Figure 6: Very good detection results of YOLOv3 even under blurred image conditions (right).

Detection of Test Objects

Neural network training was done on a desktop PC (i7 single core, OS Ubuntu 16.04) with a NVIDIA 1080Ti GPU which took approx. 8 hours (until there was no significant improvement in the loss value). The neural network architecture and weight-/configuration files were then transferred to a NVIDIA Jetson TX2 (OS: Ubuntu 16.04), where we achieved frame rates of approx. 2-3 images/sec. with a 640x480 pixel webcam. Using a 'light-weight' version of the original YOLOv3-network (named *tiny yolo*) resulted in slightly reduced detection performance but with frame rates of approx. 10 images/sec (which in both cases is sufficient for the proposed task).

The recognition of all learned label classes was excellent (Figure 6 left) – even under difficult lighting conditions or poor camera image quality (mostly due to blurring, Figure 6 right). During our tests only 2 problems could be observed:

- Whenever the hazmat labels were arranged too close to each other (Figure 7) the detection completely failed.
- If the image background (in our case in the form of a brown wooden wall) had changed during testing (compared to the training, i.e. when the hazmat labels were mounted on e.g. a white wall), a significant reduction in the recognition rate could be observed. This can clearly be traced back to a training bias, since when labelling with rectangular ROIs, especially with twisted object versions, image background structures are always necessarily 'trained' with them (Figure 8).

Figure 7: Detection completely failed when objects were arranged too close to each other.

Conclusions

Based on deep learning algorithms (YOLOv3), the presented system is able to localize and classify relevant hazard labels in the working area very robustly. This vision system enables the detection of first concrete evidence of the presence of hazardous materials regardless of the particular lighting situation (day, night, fog, etc.), over a wide range of distances and under strongly varying degrees of rotation.

The algorithm for robust real-time hazard label detection, recognition, identification and localization is running onboard a power-efficient AI computing device (NVIDIA Jetson TX2) onboard a mobile rescue robot.

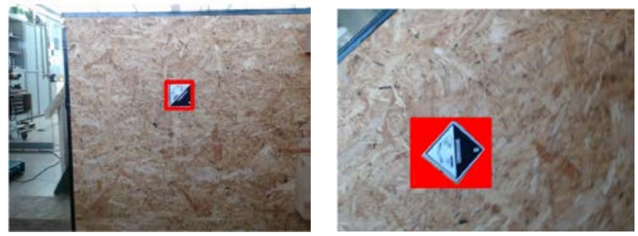

Figure 8: Image labelling for neural network training: when annotating rotated versions of an object (right), inevitably also background image structures are learned (indicated in red).

References

[1] Hiroaki Kitano et al., RoboCup Rescue: Search and Rescue in Large-Scale Disasters as Domain for Autonomous Agents Research, IEEE SMC'99, ISBN 0-7803-5731-0, Tokyo Japan, 739-743, 1999.

[2] Adam Jacoff et al., Using Competitions to Advance the Development of Standard Test Methods for Response Robots, PerMIS'12 Proceedings of the Workshop on Performance Metrics for Intelligent Systems, ISBN 978-1-4503-1126-7, Maryland, 182-189, 2012.

[3] Gill Pratt, Justin Manzo, The DARPA Robotics Challenge, IEEE Robotics & Automation Magazine, ISSN 1070-9932, 10-12, 2013.

[4] WRS, 2018. Official Homepage – World Robot Summit, Available from: http://worldrobotsummit.org/en/ Accessed: 2018-09-18

[5] Mark S. Nixon, Alberto S.Aguaro. Feature Extraction & Image Processing for Computer Vision. Academic Press, 2012

[6] S. K. P. Nikhil R. Pal. \A review on image segmentation techniques". In: Elsevier (1993)

[7] Yuanqing Lin, Fengjun Lin, Shenghou Zhu, Ming Yang. \Large-scale image classification: Fast feature extraction and SVM training". In: IEEE, CVPR (2011)

[8] Shih-Hsu Chang, Fang-Hsuan Cheng, Wen-Hsing Hsu, Guo-Zua Wu. \Fast algorithm for point pattern matching: Invariant to translation, rotation and scale changes". In: Elsevier (1995)

[9] Lowe, David G. "Object recognition from local scale-invariant features." Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.

[10] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." European conference on computer vision. Springer, Berlin, Heidelberg, 2006.

[11] Erik Hjelmås and Boon Kee Low. Face detection: A survey. Computer vision and image understanding, 83(3):236–274, 2001.

[12] Chang Huang, Haizhou Ai, Yuan Li, and Shihong Lao. Vector boosting for rotation invariant multi-view face detection. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 1, pages 446–453. IEEE, 2005.

[13] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In Proceedings of the ninth European Conference on Computer Vision, May 2006.

[14] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In Pattern Recognition, pages 297–304. Springer, 2003.

[15] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–511. IEEE, 2001.

[16] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4):399–458, 2003.

[17] Joseph Redmon, Ali Farhadi: YOLOv3 – An Incremental Improvement; https://arxiv.org/abs/1804.02767

[18] Christopher Hawley, HAZARDOUS MATERIALS: RECOGNITION AND IDENTIFICATION, online: https://www.chemsafetypro.com/Topics/USA/NFPA_704_Label_NFPA_Rating.html

[19] Yu, G., & Morel, J. M. (2009, April). A fully affine invariant image comparison method. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 1597-1600). IEEE.

[20] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., ... & Ng, A. Y. (2009, May). ROS: an open-source Robot Operating System. In ICRA workshop on open source software (Vol. 3, No. 3.2, p. 5).

Author Biography

Raimund Edlinger received his DI(FH) in sensoric (2007) and microsystems and his MSc. in Automation Engineering (2013) from the University of Applied Sciences Upper Austria. Since 2007 he has worked in the Research and Technology at the University in Wels/Austria. His work has focused on the development of mobile robots and sensor systems. He is on the board of RoboCup Rescue League as technical member and since 2018 Phd student at Graduate School Science and Technology at University of Wuerzburg.

Gerald Zauner received his PhD in applied physics from Vienna University of Technology (2005). Since then he has worked at the University of Applied Sciences Upper Austria (School of Engineering) where he is professor for signal processing (2015). His work has focused on industrial machine vision, optical metrology and non-destructive testing.

Michael Zauner received his BSc in Sensor Technology and Microsystems (2009) and his MSc. in Intelligent Sensors and Systems (2013) from the University of Applied Sciences Upper Austria. Since 2013 he is an Assistant Professor for Electrical Engineering and Robotics at the University in Wels/Austria. His work also has focusing on the development of mobile autonomous robots and the smart combination of sensor data.