

A Distributed Computing Real-Time Safety System of Collaborative Robot

Dawid Gradolewski^{1,2,*}, Dawid Maslowski², Damian Dziak¹, Bartosz Jachimczyk¹,

Siva Teja Mundlamuri^{1,2}, Chandran G. Prakash^{1,2}, Wlodek J. Kulesza^{1,3}

¹*Institute of Applied Signal Processing, Blekinge Institute of Technology,
371 79 Karlskrona, Sweden*

²*Intema Sp. z o.o.,
Siennicka 25a, 80-758 Gdansk, Poland*

³*University of Social Sciences,
9 Sienkiewicza St., 90-113 Lodz, Poland
dawid.gradolewski@bth.se*

Abstract—Robotization has become common in modern factories due to its efficiency and cost-effectiveness. Lots of robots and manipulators share their workspaces with humans what could lead to hazardous situations causing health damage or even death. This article presents a real-time safety system applying the distributed computing paradigm for a collaborative robot. The system consists of detection/sensing modules connected with a server working as decision-making system. Each configurable sensing module pre-processes vision information and then sends to the server the images cropped to new objects extracted from a background. After identifying persons from the images, the decision-making system sends a request to the robot to perform pre-defined action. In the proposed solution, there are indicated three safety zones defined by three different actions on a robot motion. As identification method, state-of-the-art of Machine Learning algorithms, the Histogram of Oriented Gradients (HOG), Viola-Jones, and You Only Look Once (YOLO), have been examined and presented. The industrial environment tests indicated that YOLOv3 algorithm outperformed other solutions in terms of identification capabilities, false positive rate and maximum latency.

Index Terms—Artificial intelligence; Collaborative robots; Neural networks; Safety system.

I. INTRODUCTION

Robots are widely used in modern industry mainly due to their effectiveness, load capabilities, precision, and repeatability. Their unique properties like fitting to the tasks, customizable smooth motion, and environmental awareness increase their applicability in difficult, monotonous tasks of various branches [1].

Nowadays, the robots and manipulators are often placed in workspaces, which they share with humans in the form of a collaborative framework, e.g., in manufacturing or assembling processes. Such machines are called “collaborative robots” or in short “co-bots”, which belong to

Intelligent Assist Devices (IADs) technology branch [2] holding \$12 billion by 2025 [3]. Especially, the small and medium size companies benefit from robotization and automatization. However, the costs are crucial for them.

Because of robot’s dynamic and close human-machine interaction, the health and safety issues become major technical challenges [4]. It has been reported that between 1984 and 2017, only in the United States, 39 accidents occurred mainly due to a lack or insufficiency of a security system. Among casualties, there were 28 robot operators, 7 maintenance workers, and 4 programmers [5].

One of the challenges of developing a highly reliable safety system is a requirement of a real-time performance in changeable surroundings without compromising co-bots’ productivity. Such a “smart” system has to detect possible dangers, identify their level, and then undertake pre-defined action. A short reaction time, reliability, and high detection and identification capabilities are crucial requirements for a suitable safety system of co-bot.

The vision-based solutions are the most effective in monitoring. Therefore, they became the core of modern safety systems. However, the amount of data needed to be processed in real-time by the system grows rapidly with the augmentation of observation zones. In most of the applications, co-bots have to work as standalone units being a part of technological processes. In such cases, a person may approach the co-bot from any side. Therefore, the safety system needs to monitor all the surroundings. With the need to cover wider safety zone, number of cameras increases, which results in a request for more computational capacity. This is the main challenge of developing modern and cost-effective safety systems, especially accessible for medium and small size companies.

The proposed solution combines the Internet of Things (IoT) technology within a distributed computing paradigm. The developed safety system consists of independent sensing/detection units, whose number and deployment could be customized for a desired observable area to match the particular co-bot’s tasks. Thanks to the designed Graphical User Interface (GUI), each module can be easily

Manuscript received 17 October, 2019; accepted 23 February, 2020.

The model was developed in *Intema Sp. z o.o.* (Siennicka 25a, 80-758 Gdansk, Poland) and is currently under the evaluation. This research was funded under Grant (No. POIR.01.01.01-00-0074/17) from the National Centre for Research and Development of Poland.

re-configured. The data from up to six sensing and detecting units are sent and then processed in real-time by a capable Graphical Processing Unit (GPU) server what makes the proposed solution cost-effective.

II. SURVEY OF RELATED WORKS

Detection of human presence within the robot vicinity is a complex and demanding task. Numerous attempts have been made to develop a robust reliable detection system. This survey of related works focuses on the scope of modern technologies and sensors for both the co-bots, manipulators and mobile platforms. Apart of available hardware solutions, the state of the art of methods and algorithms used is presented.

A. Technologies and Sensors

The conventional approach to a safe Human-Robot Collaboration (HRC) is a pre-collision strategy commonly applied in the manufacturing environment [6]–[8]. To ensure the safe co-operation between humans and robots, a human detection [9], [10], human pose estimation [11], and obstacle avoidance [12] are typically needed. Many solutions, which have been already implemented on co-bots safety systems apply depth sensors [7], [8], [11], vision systems [13], LiDAR [14], [15] or RADAR [16] technologies. In order to detect the physical contact between human and robots, touch sensors are useful [17], [18].

Nevertheless, due to the information relevancy about the object's morphology, a vision technology becomes the most popular approach in safe HRC. In general, there are two kinds of vision systems used in safe HRC: single camera [19], [20] and stereo vision [7], [21]–[23]. While the standalone, however, the more expensive and demanding stereo vision technology allows estimating the 3D coordinates of the detected object. The single camera solutions require additional sensors or systems for proximity measure.

Due to required localization precision, the stereo vision uses the High-Resolution (HR) cameras [24]–[26], while the single camera system does not demand expensive HR solutions to detect the object and regular industrial tools are sufficient [27]–[31].

B. Object Localization Techniques for Vision Systems

The need for the localization of detected object in real-time usually leads to a fusion of methods or algorithms, which combine the data from various sensors. One of the first approaches used a single camera vision and ultrasonic sensing for obstacle detection [27]. The system distinguishes between the stationary and moving objects. Stationary objects are detected using the vision system by means of edge detection, whereas the moving objects are detected using the ultrasonic sensing. In [12], authors present the pre-collision strategy by exteroceptive sensing framework based on multiple Kinects deployed in the working cell. A similar approach is applied in the solution proposed by Mohammed, Schmidt, and Wang [8], where robot's virtual 3D models are associated with human operators using a series of depth and vision sensing units for online collision detection and avoidance in an augmented environment.

Other vision-based solutions use stereo vision for identification and localization of detected object. Ebert *et al.* [24], [25] developed a collision detection system based on a set of images taken simultaneously by several cameras. Here, the obstacle shape is reconstructed by the look-up-table-based fusion algorithm and the 3D images. The real-time human tracking system proposed by Petrovic *et al.* uses stereo vision and the Kalman filter for object tracking [26]. This system enables the object detection with 96 % reliability of classification performance. The reduced computational complexity on the determined region of interest allow the robust real-time performance.

C. Detection, Identification, and Classification algorithms

The object detection algorithm is the core of each vision safety system either it is based on single camera or stereo vision approach. However, the vision-based real-time object detection algorithms that may be implemented in safety system have a relatively short history. The main challenges of the systems are the computational complexity and diversity of cases (humans, environmental conditions).

The Viola-Jones object detection released in 2001 organizes Haar-like features in so-called "classifier cascade" what simplifies the task of Haar classification and ensures better performance [33]. The Histogram of Oriented Gradients (HOG) released in 2005 is a feature descriptor and often is combined with a support vector machine for identification [29]. Both solutions are very fast. The approaches that uses Haar-like feature extraction (Viola-Jones) for face detection allows identifying objects at rate of an half a second [34]. Redmon *et al.* propose the safety system working in real-time on 45 Frames Per Second (FPS) with 25 ms latency based on the fast-single layered Neural Network (NN). However, to reduce the computational complexity, the resolution of the image has to be downsized to 448 px×448 px [28]. Although the system localization errors are evenly distributed up to 19 %, it shows better performance than other object detection algorithms like HoG [29], Harr [30], Deformable Parts Model (DPM) [31], Region-based Convolutional Network method (R-CNN) [32], and Speeded Up Robust Features [35]–[39].

With the grow of GPU capabilities, the Artificial Intelligence-based systems (AI) become more and more powerful. Since 2012, when Krizhevskys' Convolution Neural Network (CNN) [40] has won the ImageNet competition, NN becomes the new standard for image classification. AI outperforms others approaches in human recognition and prediction of human activity within the co-bots' workspace. An example is human activity prediction proposed by Ding *et al.* in [41]. Another AI-based solution is a context-aware safety system for real-time HRC in terms of collision sensing and path planning. The essential system's feature is context-aware human pose recognition, which was carried out by the CNN [42]–[45].

The deep learning became a breakthrough in computer vision and object detection. Girshick *et al.* propose the Region-CNN (R-CNN) algorithm, which selectively searches regions in the image for a further classification purpose [46]. Another object detection algorithm presented in [47] differs from the region-based solutions. However, the

best results are obtained with the You Only Look Once (YOLO) Deep Neural Network (DNN) [28].

III. PROBLEM STATEMENT, MAIN OBJECTIVES AND CONTRIBUTIONS

As the survey of related works shows, none of the existing safety systems enables the real-time robust detection and identification of humans or other objects approaching the operational zone of co-bots. So far, the best results are achieved from the vision-based systems. However, a needed number of high-resolution cameras to cover the entire surroundings of the machine results in a demand of high computational capability, which challenges the system real-time performance. Furthermore, an efficient implementation of the identification algorithms requires a large number of GPU platforms, which further increases the solution's price. To overcome these problems, designers are forced to reduce the image resolution what affects the system performance in terms of its robustness and accuracy.

The main objective of this paper is to find a reliable cost-effective real-time solution of a safety system of co-bot for detecting, identifying, and localizing humans approaching the operational zone, and providing suitable information to the co-bot control system. The system reliability in changeable surroundings without compromising machine productivity is crucial.

The proposed safety system has a modular structure consisting of independent sensing/detection units, whose number and deployment could be customized for the desired observable area and co-bot's tasks. Each sensing/detection unit is composed of a HD vision camera, an ultrasound

sensor directable by a stepper motor, and a supplementary controller. The included microcontroller aims on preprocessing the HD images for motion detection. To fulfill the real-time request, the authors propose to apply the distributed computing concept. Therefore, the GPU server with an implemented AI algorithm analyses the images of the detected objects, localizes, identifies them, and then classifies potential threats.

The solution is implemented on an off-the-shelf component of the Raspberry Pi microcontroller for processing and controlling unit for motion detection and proximity estimation. The motions detection is based on a HD single camera and proximity estimation ultrasound sensors. The NN-based decision-making system for identification and classification is implemented on GPU with Linux machine. The system was verified on Braccio Tinkerkit robot [48].

IV. SYSTEM DESIGN

In this research, the User-Driven-Design (UDD) methodology is used [48], where the design process is systemized and both stakeholders and future users are involved at each design stage. The defined general and itemized functionalities are summarized in Table I. The demands that the system should be easy to install, operate and maintain, along with low cost and size, are classified as overall constraints. Both general and itemized functionalities, particular constraints, and possible technologies and algorithms are summarized in Table I. The selected most suitable technologies and algorithms are indicated in bold.

TABLE I. GENERAL AND ITEMIZED FUNCTIONALITIES OF DEVELOPED SAFETY SYSTEM.

Functionalities		Particular Constrains	Possible Technologies and Algorithms
General	Itemized		
Safety System	Reliability, Cost-effectiveness	Real time performance, Distance range < 5 m	RFID, RTLS, Vision, Ultrasound, Distributed computation , CUDA, GPU
Object detection	Image segmentation	Detected object size > 1000 px, Computation rate < 10 FPS	Vision system, Ultrasound sensor, Radar, Lidar, Background subtractor, Cropping
Detection Classification	Human, another objects	Reliability > 98 %, Simultaneously up to 4 people/objects.	Neural networks, Genetic algorithms, Decision Trees Haar, Hog, YOLO
Object localization	Distance estimation	Latency < 200 ms Localization uncertainty < 10 %	Vision, Radar, Lidar, Ultrasound sensors , Stereo vision, Triangulation, Beamformer, Pulse-echo technique
Threat classification	Classification zones	Safe Zone (2 m–5 m), Unsafe Zone (1 m–2 m), Hazardous Zone < 1 m.	Machine Learning, Distributed computation , Decision trees, SVM, kNN, Deep NN
Co-bot motion control	Speed control,	For unsafe zone speed reduced by half	Programmable Safety Control , Application Programming.
	Stop control	For hazardous zone	
Interfacing	Connection reliability	Ping < 100 ms, Dropped signals rate < 1 per day	Bluetooth, Wi-Fi, ETHERNET, TCP/IP, UDD

The request was that the designed safety system should be able to detect and track simultaneously up to four moving objects from at least five meters distance from the robot. The system has to ensure the real-time performance what means 1 s response time from the object detection to robot reaction, which is necessary to smoothly change the co-bots' path.

From 5 m distance, the safety system has to detect movements caused by objects of at least 350 cm² what corresponds to size of the human head. System performance needs to be highly reliable in a changeable light environment from at least 100 lumens. The desired reliability of human identification shall exceed 98 %. Localization uncertainty

sufficient for proximity sensing subsystem needs to be 10 % and the maximum latency is 200 ms.

The system has to perform specified actions when a human is approaching to two of the three defined safety zones. In *Safe Zone*, which is situated between 2 m and 5 m from the robot, the object should be only tracked, non-action is required. In *Unsafe Zone*, which is situated between 2 m and 1 m, the system should reduce the working speed by half. In *Hazardous Zone*, which is less than 1 m from a robot, all robot's movements have to stop.

The determined functionalities and constrains were evaluated as technically accomplishable and feasible. The possible technologies and algorithms presented in Table I indicate that there are several ways to solve the problem. However, the overall constrains of small size and low cost limit the use of certain solutions.

Finally, the modular system based on distributed computation paradigm using a microcontroller Raspberry Pi and only one GPU server is selected. Sensing/detection modules are responsible for motion detection and distance measurement, and the CPU server's main tasks are to identify the selected object and to make decision. Each sensing/detection module is equipped with a camera, microcontroller, and ultrasound distance sensor that constitute a low cost solution, which does not require substantial computation power.

Among the excluded solutions, there is, e.g., Radio-Frequency Identification (RFID) technology, which requires tagging of all tracked objects what is not acceptable since someone without a tag can intrude co-bots' operating zone. Other technologies, such as Compute Unified Device Architecture (CUDA), in turn require a significant amount of computing power and/or are high price solutions.

The chosen solution of object detection, localization, and classification is based on vision signal processing by image background subtraction, cropping, and Machine Learning algorithms. The ultrasound distance measurement applies pulse-echo algorithms. All algorithms used are characterized by their simplicity, reliability, and low computational complexity.

For threat classification, DNN is chosen, which analyses vision and distance data from the modules. The DNN is chosen due to its high reliability. The co-bot motion is controlled by means of Programmable Safety Control (PSC), the software built in co-bot. As a communication medium, the TCP/IP and Ethernet are used because of their high speed, reliability, interference immunity, and safety of data transfer.

V. SYSTEM DEVELOPMENT

A. Architecture of the System

The architecture of the system is presented in the Fig. 1. The system is designed based on IoT concept, where a user has easy access to each subsystem through the GUI. All data of the previous detections and regarding system performance are stored on one database.

The safety system is composed of sensing/detection modules, decision-making system, and robot, which are connected via Ethernet. Data processing is decentralized,

where the decision-making system deals with the preprocessed data from the sensing/detection modules.

Each sensing/detection module is equipped with the vision and ultrasound system, respectively. The main task of vision system is motion detection, whereas the ultrasound system is responsible for distance measurement. The core detection algorithm is implemented on the microcontroller, which is of-the-shelf Raspberry Pi. When object motion is detected, the ultrasound system is directed towards the object to measure the accurate distance. The view of the designed single sensing/detection module is presented in Fig. 2.

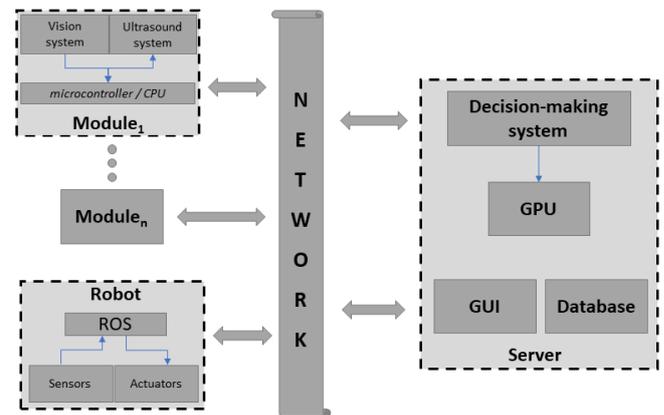


Fig. 1. Architecture chart of a designed safety system.

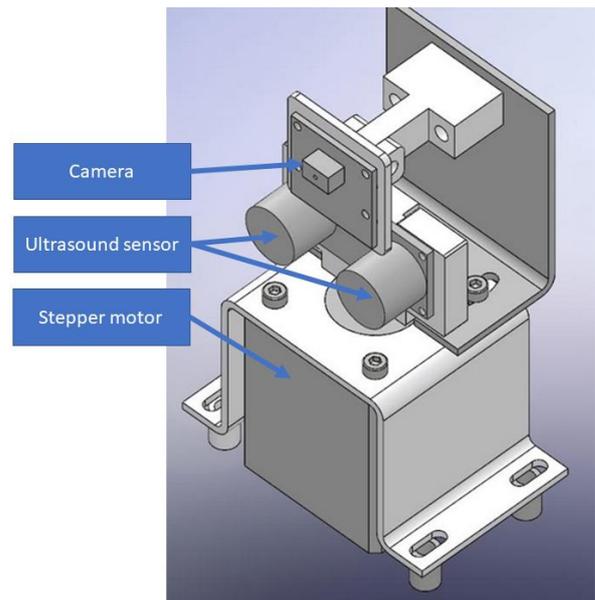


Fig. 2. Single sensing/detection module composed of camera, ultrasonic sensor, and stepper motor.

The number and configuration of the sensing/detection modules depend on customer's request and co-bots' vicinity. To cover all the surroundings, up to 6 modules are needed. The module configuration schema is presented in the Fig. 3, where *Zone A* (depicted in black) is the co-bot working area, *Zone B* (depicted in red) is the hazardous zone, where a direct contact between machine and human can occur, and to prevent it the co-bot have to be stopped. *Zone C* (depicted in yellow) is the warning zone, where a detected object has to be observed and the co-bot motion speed needs to be reduced by half. *Zone D* (depicted in green), which is the detection "safe" zone (2 m–5 m), where the preliminary object detection and identification are requested.

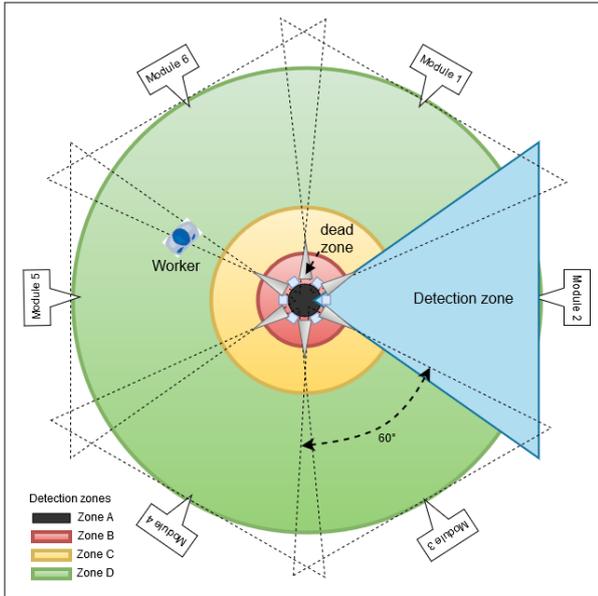


Fig. 3. Safety system configuration schema of sensing/detection modules.

The Field of View (FoV) of the single sensing/detection module covers a 60° circle section of a range up to 5 m. One can observe that there exists a small dead zone between neighboring sensing/detection modules. However, its size is smaller than human's head, and therefore it can be neglected

since it does not affect the system performance in terms of its security. The dead zone is in the shape of an isosceles triangle with a base of 25 cm and a height of 50 cm.

The cropped images consisting of detected objects are combined with information about their location, and then sent to the decision-making system for identification. The decision-making system is implemented on a GPU. After a positive identification, the system sends to the Robot Operating System (ROS) the warning about a presence of an identified human within a defined detection zone. Meanwhile, the data are backed up at a database with the images and detailed information about the event. Status of each sensing/detection module is available online in the TCP/IP and Json standards. This information is accessible to the user through the GUI.

The main task of the decision-making system is to identify the detected object. Thanks to the distributed computation approach, the decision-making system is able to handle data from up to six independent sensing/detection modules.

B. Model of the System

The flowchart of co-bot safety system algorithm is presented in Fig. 4, where the sensing/detection modules are implemented on embedded microcontrollers, and the decision-making system is executable on the servers' GPU.

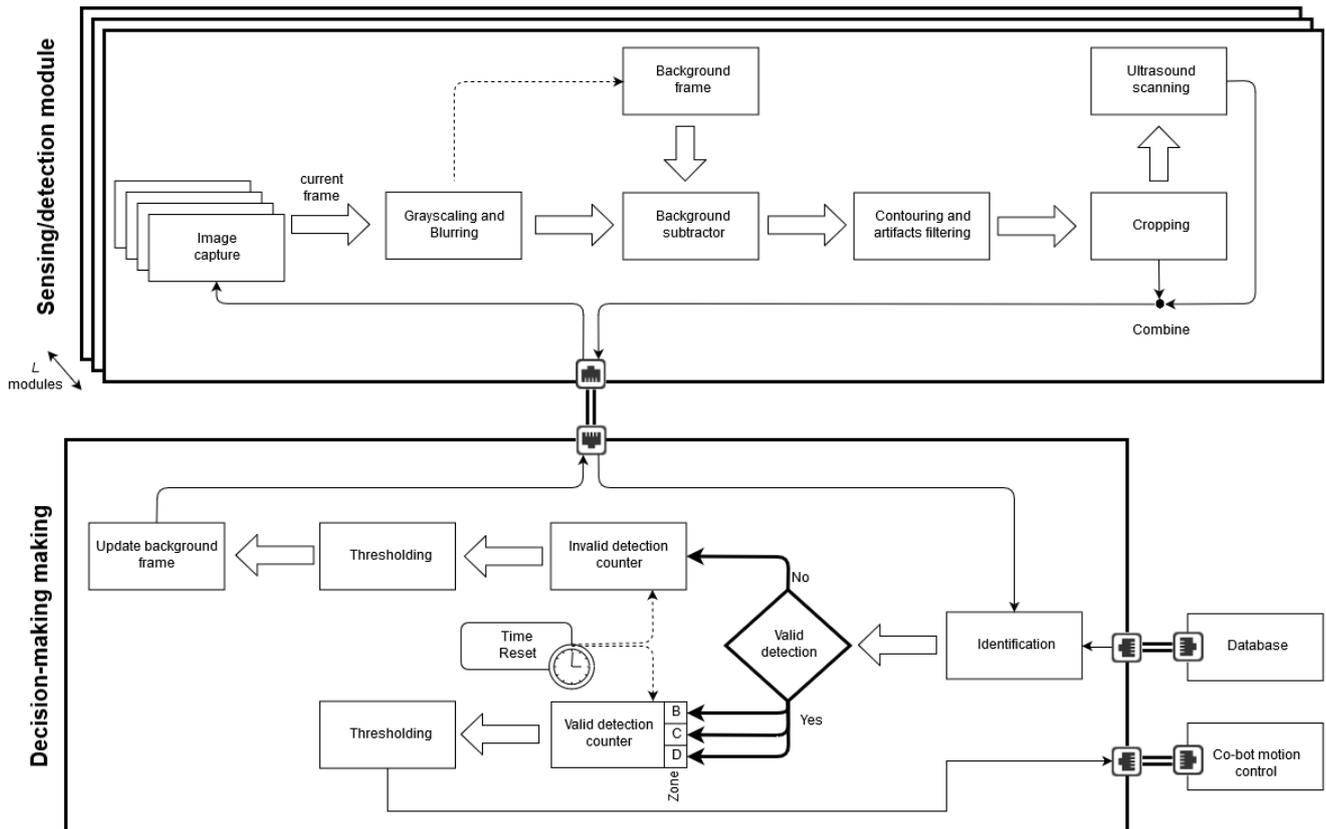


Fig. 4. The co-bot safety system flowchart.

The algorithm of sensing/detection module is based on the motion detection principles (Fig. 5). First, the background and current images are converted into greyscale, where 0 represents black color and 255 is white one. After blurring, the absolute differences between corresponding pixels of the two images are calculated. Next, the differentiated image is contoured and filtered by thresholding, which removes

artefacts caused by the light changes, etc. As a result, an image binary matrix is obtained, where 0 stands for detected object and 1 means background. Later, the area of the detected object is calculated from the object contour. If the area is bigger than assumed one, the cropped frames are defined and applied on the original color image.

Once the object is detected and its direction is localized,

the ultrasound sensor is directed towards the object to measure its distance from the co-bot. Finally, the object's cropped image together with its localization data are sent to the decision-making system to decide whether to perform a suitable action. The cropped frames together with distance data are fed to a Machine Learning algorithm to classify if the detected object is a *person*. The algorithms tested for this application were: the Histogram of Oriented Gradients (HOG [29], Viola-Jones [33], and YOLO [28].

To improve the safety system robustness, valid and

invalid detection counters are included. The goal of this solution is to reduce the overall false detections introduced by, e.g., isolated detections at the border of detection area or false detections caused by lighting or background changes. When a number of invalid detections is higher than assumed threshold one, then it is supposed that lighting or background had changed and a new background frame would be triggered. Also, a number of valid detections needs to exceed a threshold number to trigger an action from the robot control system.

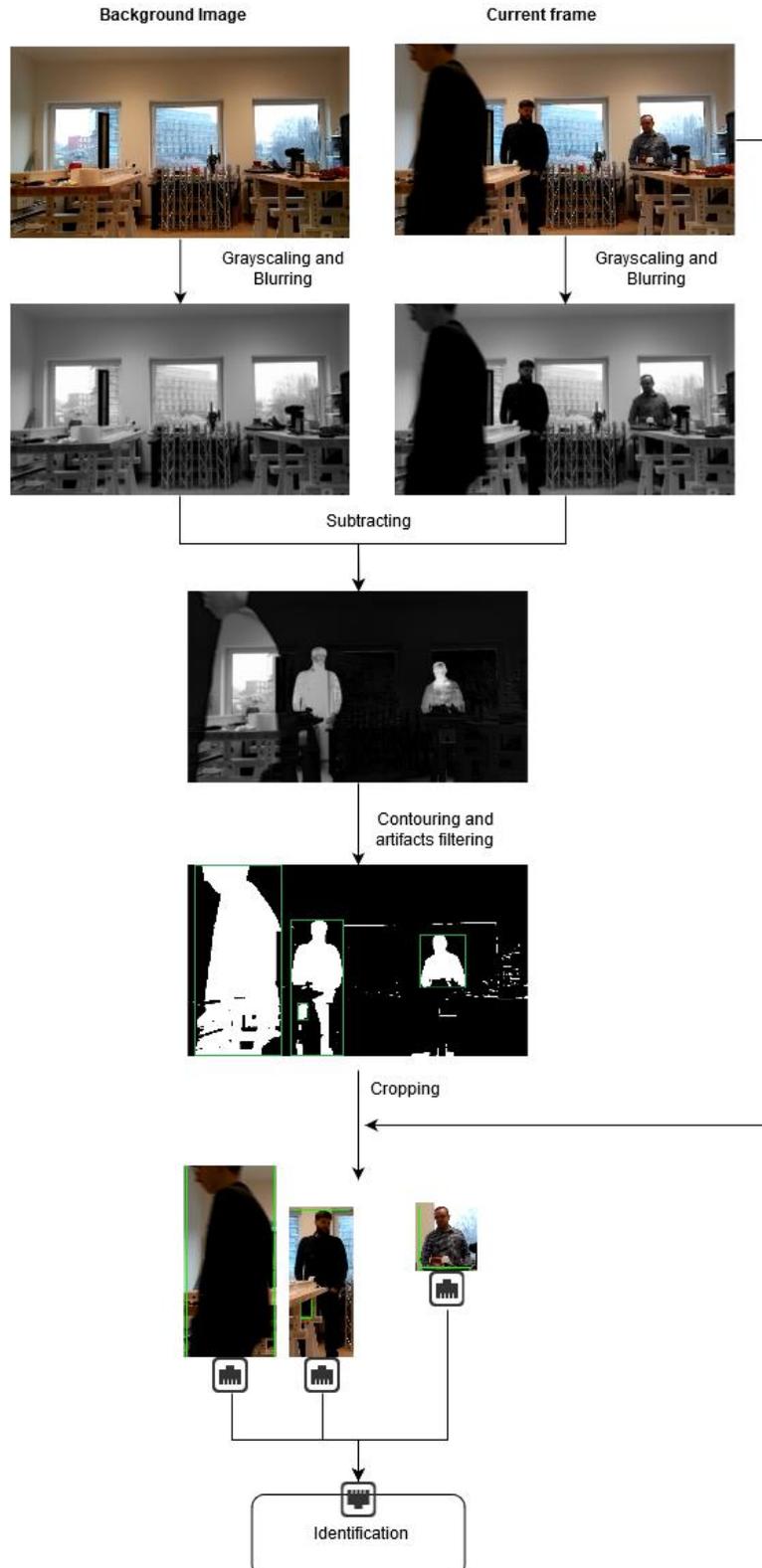


Fig. 5. Image processing part of co-bot safety system implemented on sensing/detection module.

C. Prototype of the System

A picture of two sensing/detection modules is presented in Fig. 6. In Fig. 7, the designed safety system covering 360° FoV observation area embedded in the co-bot is depicted. All the sensing/detection modules are installed on the co-bot base. Each module includes a fixed 8 Mpx camera [49], a 30 kHz ultrasound sensor of range from 2 cm to 5 m

[50], a bipolar four-cord stepper motor with step resolution 0.9° [51], Raspberry Pi 3 B+ with 1.4 GHz CPU, 1 GB RAM, ARM architecture, and 16 GB micro memory card [52]. The 5 V DC, 2 A power supply for all the components is also applied [53]. The decision-making system comprises an industrial Dell workstation with I7-8750H six cores 3.1 GHz CPU, 32 GB RAM, and NVIDIA GeForce GTX 1050 Ti GPU with 768 CUDA cores and 4 GB RAM [54].

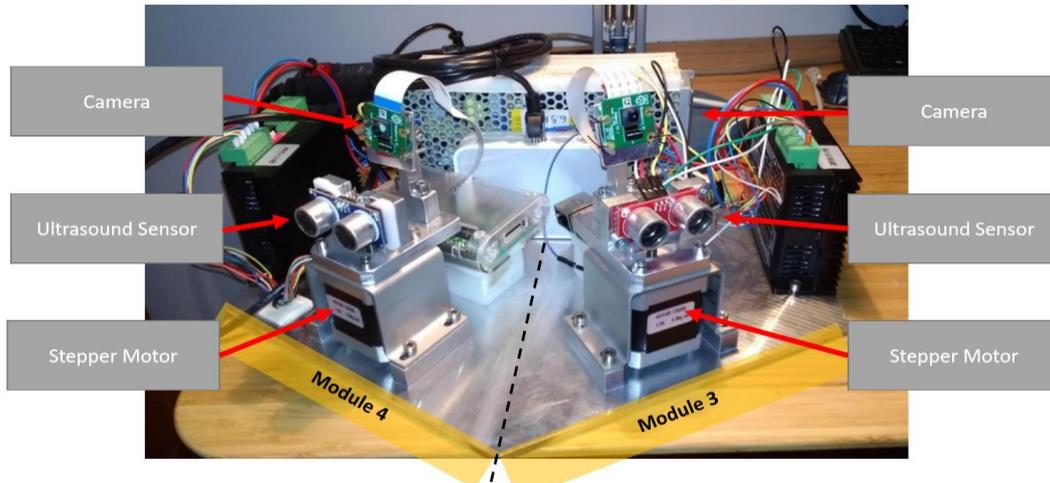


Fig. 6. Two-module sensing prototype of the safety system.

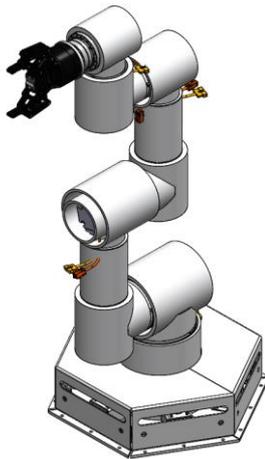


Fig. 7. Model of the robot with proposed safety system installed in its base.

The parameters of image pre-processing algorithms were chosen heuristically based on tests performed in various light conditions and on different backgrounds. The number of persons appearing in the detection area, their appearance and growth also varied. Based on the test results, the blurring frame size was chosen for 11 px. A smaller frame value would increase the contour precision, which on other hand would cause more false positive detections. A greater blurring frame could cause division of the detected object into smaller parts. The threshold level for calculation of an image binary matrix used for contouring was chosen for 50. To be further processed, the detected object had to be greater than 500 px, what corresponds to the size of human head recorded from distance of 6 m.

The threshold of valid detection counter depends on the sampling rate in terms of Frame Per Seconds (FPS), and was selected analytically. Since the maximum latency required by the users and stakeholders was 0.2 s, then to classify a

detection as valid the counter needs to be more than 50 % of all possible detection in a considered period of time. Therefore, the value of threshold (N_{VD} within 0.2 s) time, which is triggered by the first counting, could be calculated using the following formula

$$N_{VD} = FPS \times 0.2 / 2. \quad (1)$$

If the estimated sampling rate is equal to 10 FPS, then the object had to be identified at least twice within 0.2 s to trigger any action of the system. The FPS rate may vary due to CPU overload or temperature change inside the module. Therefore, the current FPS value of each sensing/detection module is estimated and sent to the server for each 1 min interval.

A need for background image update does not depend on the latency request. The invalid detection counter has been selected analytically as 30 false detections within 60 s triggered by the first false detection.

The tree identification algorithms, Hog, Viola-Jones, and Yolo have been selected for testing. These state-of-the-art algorithms have been examined and compared in terms of the detection efficiency, false positive rate, maximum latency measured as a delay from an instant of object detection to its identification, and computational complexity. For this purpose, the 120 s test was done at the same light conditions and with the same image resolution of 1360 px×768 px. One person walked with normal speed with random path forward and backward in the room at the distance between 0.5 m and about 4.0 m was monitored.

The detection efficiency was calculated as a ratio of a number of valid frames - correctly identified - to the number of all frames. The false positive (FP) is an identification of a *non-person* object as a *person*, and the false positive rate (FP_R) is calculated as the number of false positive

identifications related to the number of false identifications, which is the sum of numbers of false positive identifications and true negative (TN) identifications

$$FP_r = \frac{FP}{FT + TN}. \quad (2)$$

The illustrative examples of false detection are presented in Fig. 8. The latency was measured as the delay between the detected appearance of the object in the zone and its identification by decision-making system. Computational complexity is expressed in terms of sampling rate measure as FPS. Table II summarizes the test results.

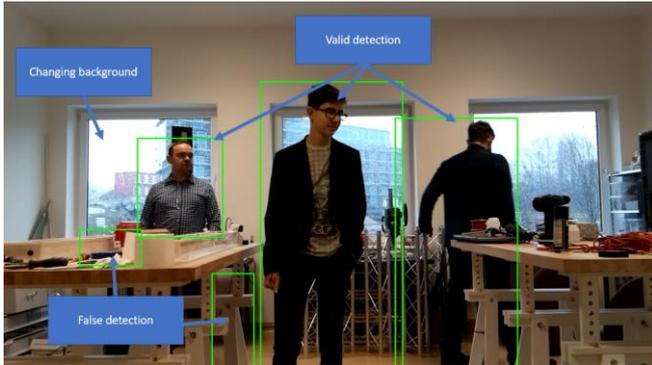


Fig. 8. Example detection of a sensing/detection module.

TABLE II. COMPARISON OF THE IDENTIFICATION ALGORITHM.

Algorithm	FP_r	Detection efficiency	Maximum Latency [ms]	FPS
HoG	0.79	70 %	201	11
Viola Jones	0.87	77 %	191	11
YOLO v3	0	99 %	151	10

Overall, the Yolo v3 outperforms other algorithms in terms of detection efficiency, false positive rate, and maximum latency. The FPS rate of all three algorithms are very similar and met the user's requirement. Therefore, the YOLO v3 algorithm was implemented in the final version of the prototype, which was validated, and it is described in the following chapter.

To select a suitable camera, an impact of the image resolution on the system latency and FPS was estimated. A 5 min test with one person moving was done. The results are presented in Table III. They show that user's requirements of latency to be less than 200 ms and computation rate higher than 10 FPS are met by a HD camera.

TABLE III. IMPACT OF THE RESOLUTION ON THE COMPUTATION LATENCY AND FPS.

Resolution [px×px]	FPS	Latency [ms]
VGA	30	101
SVGA	30	102
WXGA	16	143
HD	12	151
HD+	10	243
FHD	5	311

VI. RESULTS AND DISCUSSION

The system validation scenario concerns assessments of

the following system features: (i) ability of multi-detection, (ii) detection reliability, (iii) latency, (iii) FPS performance, and (iv) real-time performance. During the validation test, from one to three persons walked into the detection area with typical speed and with random paths forward and backward at the distance from the robot between 0.5 m and about 4.0 m. The test was performed on a single sensing/detection module. The light condition and the obstacles in background varied during the test. The 1 min test was performed when the sensing/detection module captured 601 images. An example of the test frame is presented in the Fig. 8, where in addition to the valid detections of people, some invalid detections are also visible. However, the selected YOLO algorithm managed to identify them as invalid.

From the test results, one can conclude that simultaneous detection of several objects present on images is possible, and even filtering of invalid detections is possible in real-time.

The test proves that the safety system fulfils stakeholders' requirements in terms of detection reliability, latency, computation capacity, and real-time performance request. The calculated sampling rate was 10 FPS. The maximum latency defined as a delay between the instant when the object was detected, and its first identification was measured as 161 ms.

The measured detection efficiency defined as a ratio of a number of valid frames - correctly identified - to the number of all frames was 99 %. Only for 15 cases (0.73 %) of over 2060 detected objects, the system combined two persons into one frame. In 9 % of all the cases, the image preprocessing algorithms split persons' body into parts. However, in all these cases, YOLO was still able to correctly identify the human.

Another test was dedicated to estimate the accuracy of the localization measurement of detected object. The system was tested in the laboratory of area of 160 m² with the lighting condition of 500 lux. One human at a time walked into the observation area and approached the points marked on the ground as shown in Fig 9. The marked points were distributed within three different angle ranges (10 °÷15 °, 30 °÷35 °, and 45 °÷50 °) and placed in four distances from the system (100 cm, 220 cm, 350 cm, and 470 cm). The system was tested on 30 different people, and over 300 localizations were measured. Test results are presented in the Table IV.

TABLE IV. IMPACT OF THE RESOLUTION ON THE COMPUTATION LATENCY AND FPS.

Distance [cm]	Mean [cm]	Uncertainty maximum [cm]	Uncertainty minimum [cm]	SD [cm]
100	-2.4	3	-8	2.6
220	10.6	34	-6	10.5
330	-1.2	59	-65	25.5
470	-14.3	94	-95	49.5

For different distances, the measurement uncertainty varies from 3 % up to 20 %, the mean value of the localization uncertainty is less than 6 %. The Standard Deviation (SD) of the measurement uncertainty at different

distances varies from 3 % to 11 %. Due to the used valid detection counter, at least two valid detections and distance measurements are done before sending a request for the co-bot action. Therefore, the uncertainty of distance estimation of the detected object is a half of a single measurement. Table IV summarizes the test results.

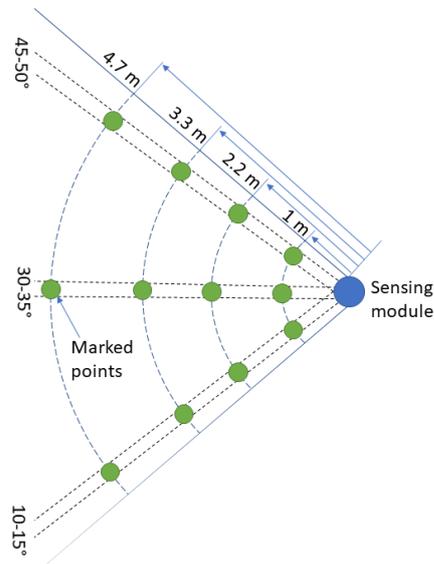


Fig. 9. Configuration of the marker points during the test.

VII. CONCLUSIONS AND FUTURE WORK

This article shows an opportunity to design the cost-effective reliable real-time safety system of co-bot, which is able to detect, identify, and localize human multi-object approaching the co-bot operational zone, and then providing a decision about a suitable action of the co-bot control system.

The systematic developing approach of User-Driven-Design (UDD) methodology facilitated a developing of the system, which met all requirements and constrains defined by the future users and stakeholders.

The system modular construction embedded into distributed computing paradigm based on the shelf-components ensures the high detection efficiency. Moreover, the modularity of the system, which included distributed computing, made it useful even in demanding industrial environment.

The division of observation area into three zones ensures system high effectiveness without compromising the productiveness of co-bot.

Conducted tests show that for the designed safety system, YOLO v3, the ML algorithm, performs best for person identification. The person detection efficiency achieved by the algorithm reached 99 %. In demanding industrial environment with many obstacles, less than 1 % of detections combined two persons in one frame, and in 9 % of cases, the detection module split on the person image in parts. Nevertheless, the YOLO algorithm managed to identify the human in all disturbed cases.

The achieved detection and localization range of the safety system is 5 m. The system works on various light conditions and on changeable background. The calculated FPS rate for the used HD camera is at least 10 FPS, which

fulfils the real-time performance request without compromising the detection efficiency and robustness. The latency measured as a delay between detection of object appearance in the detection area to its identification by the system does not exceed 0.2 s. The localization uncertainty at the maximum distance of 4 m does not exceed 10 %, which ensures action timing if a person would approach the co-bot operational zone.

The proposed safety system was implemented as an integrated part of the designed co-bot. However, it may be also used as a standalone system for every and each robot, manipulator or even a mobile platform [55].

In the future, to enhance the system performance in terms of used resolution, latency and FPS, other microcontrollers like NVIDIA Jetson Nano or Jetson TX2 with on board CUDA could be tested [56]. Their better computational capabilities shall help to extend the system applicability.

The plans for system enhancement also include full tracking of a person and taking action only when the person is actually heading to the co-bot operational zone.

Further system development assumes a use of advanced AI in decision-making system, such as DNN or decision trees.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] J. Faneuff and J. Follett, "Human-robot collaboration", in *Designing for Collaborative Robotics*. O'Reilly Media, Inc., 2016.
- [2] Intelligent Assist Devices (IADs), 20 Jun. 2019. [Online]. Available: <https://www.assemblymag.com/articles/82674-intelligent-assist-devices>
- [3] J. Kite-Powell, "This new robotic avatar arm uses real time haptics", *Forbes*, 26 Feb. 2018. [Online]. Available: <https://www.forbes.com/sites/jenniferhicks/2017/11/29/this-new-robotic-avatar-arm-uses-real-time-haptics/>
- [4] M. Vasic and A. Billard, "Safety issues in human-robot interactions", in *Proc. of 2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 197–204. DOI: 10.1109/ICRA.2013.6630576.
- [5] Accident Search Results Page, Occupational Safety and Health Administration, 2018. [Online]. Available: <https://www.osha-slc.gov/accident-search/>
- [6] N. Najmaei and M. R. Kermani, "Prediction-based reactive control strategy for human-robot interactions", in *Proc. of 2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 3434–3439. DOI: 10.1109/ROBOT.2010.5509179.
- [7] F. Flacco, T. Kröger, A. D. Luca, and O. Khatib, "A depth space approach to human-robot collision avoidance", in *Proc. of 2012 IEEE International Conference on Robotics and Automation*, 2012, pp. 338–345. DOI: 10.1109/ICRA.2012.6225245.
- [8] A. Mohammed, B. Schmidt, and L. Wang, "Active collision avoidance for human-robot collaboration driven by vision sensors", *Int. J. Comput. Integr. Manuf.*, vol. 30, no 9, pp. 970–980, Sep. 2017. DOI: 10.1080/0951192X.2016.1268269.
- [9] J. Krüger, T. K. Lien, and A. Verl, "Cooperation of human and machines in assembly lines", *CIRP Ann.*, vol. 58, no 2, pp. 628–646, Jan. 2009. DOI: 10.1016/j.cirp.2009.09.009.
- [10] A. Koschan, C. Cheng, M. Abidi, C. Chen, and D. Page, "Tracking a moving object with real-time obstacle avoidance", *Ind. Robot Int. J. Robot. Res. Appl.*, vol. 33, no 6, pp. 460–468, Nov. 2006. DOI: 10.1108/01439910610705635.
- [11] F. Flacco and A. D. Luca, "Safe physical human-robot collaboration", in *Proc. of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 2072–2072. DOI: 10.1109/IROS.2013.6696635.
- [12] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward safe human robot collaboration by using multiple kinects based real-time human tracking", *J. Comput. Inf. Sci. Eng.*, vol. 14, no. 1, p. 011006,

2014. DOI: 10.1115/1.4025810.
- [13] S. M. M. Rahman, Z. Liao, L. Jiang, and Y. Wang, "A regret-based autonomy allocation scheme for human-robot shared vision systems in collaborative assembly in manufacturing", in *Proc. of 2016 IEEE International Conference on Automation Science and Engineering (CASE)*, 2016, pp. 897–902. DOI: 10.1109/COASE.2016.7743497.
- [14] M. Zhao, S. Stasinopoulos, and Y. Yu, "Obstacle detection and avoidance for autonomous bicycles", in *Proc. of 2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, 2017, pp. 1310–1315. DOI: 10.1109/COASE.2017.8256281.
- [15] P. Long, C. Chevallereau, D. Chablat, and A. Girin, "An industrial security system for human-robot coexistence", *Ind. Robot Int. J. Robot. Res. Appl.*, vol. 45, no. 2, pp. 220–226, Dec. 2017. DOI: 10.1108/IR-09-2017-0165.
- [16] M. Geiger and C. Waldschmidt, "160-GHz radar proximity sensor with distributed and flexible antennas for collaborative robots", *IEEE Access*, vol. 7, pp. 14977–14984, 2019. DOI: 10.1109/ACCESS.2019.2891909.
- [17] A. D. Luca and F. Flacco, "Integrated control for pHRI: Collision avoidance, detection, reaction and collaboration", in *Proc. of 2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 288–295. DOI: 10.1109/BioRob.2012.6290917.
- [18] G. Doisy, "Sensorless collision detection and control by physical interaction for wheeled mobile robots", in *Proc. of 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012, pp. 121–122. DOI: 10.1145/2157689.2157715.
- [19] T. Yamamoto, Y. Yamada, M. Onishi, and Y. Nakabo, "A 2D safety vision system for human-robot collaborative work environments based upon the safety preservation design policy", in *Proc. of 2011 IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 2049–2054. DOI: 10.1109/ROBIO.2011.6181593.
- [20] A. Alhamwi, B. Vandepoortale, and J. Piat, "Real time vision system for obstacle detection and localization on FPGA", in *Proc. of International Conference on Computer Vision Systems*, pp. 80–90. DOI: 10.1007/978-3-319-20904-3_8.
- [21] T. K. S. Cheung and K. T. Woo, "Human tracking in crowded environment with stereo cameras", in *Proc. of 2011 17th International Conference on Digital Signal Processing (DSP)*, 2011, pp. 1–6. DOI: 10.1109/ICDSP.2011.6004902.
- [22] M. A. Mohammed, A. I. Melhum, and F. A. Kochery, "Object distance measurement by stereo vision", *Int. Journal Sci. Appl. Inf. Technol.* (Special Issue of ICCTE 2013), vol. 2, no. 2, pp. 5–8, 2013.
- [23] A. Stanoev, N. Audinet, S. Tancock, and N. Dahnoun, "Real-time stereo vision for collision detection on autonomous UAVs", in *Proc. of 2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, pp. 1–6. DOI: 10.1109/IST.2017.8261524.
- [24] D. M. Ebert and D. D. Henrich, "Safe human-robot-cooperation: Image-based collision detection for industrial robots", in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002, vol. 2, pp. 1826–1831. DOI: 10.1109/IRDS.2002.1044021.
- [25] D. Ebert and D. Henrich, "Safe human-robot-cooperation: Problem analysis, system concept and fast sensor fusion", in *Proc. of International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2001, pp. 239–244. DOI: 10.1109/MFI.2001.1013541.
- [26] E. Petrović, A. Leu, D. Ristić-Durrant, and V. Nikolić, "Stereo vision-based human tracking for robotic follower", *Int. J. Adv. Robot. Syst.*, vol. 10, no. 5, p. 230, 2013. DOI: 10.5772/56124.
- [27] I. Ohya, A. Kosaka, and A. Kak, "Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing", *IEEE Trans. Robot. Autom.*, vol. 14, no. 6, pp. 969–978, Dec. 1998. DOI: 10.1109/70.736780.
- [28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", in *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.
- [29] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in *Proc. of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893. DOI: 10.1109/CVPR.2005.177.
- [30] C. H. Setjo, B. Achmad, and Faridah, "Thermal image human detection using Haar-cascade classifier", in *Proc. of 2017 7th International Annual Engineering Seminar (InAES)*, 2017, pp. 1–6. DOI: 10.1109/INAES.2017.8068554.
- [31] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010. DOI: 10.1109/TPAMI.2009.167.
- [32] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection", *IEEE Trans. Multimed.*, no. 99, p. 1–1, 2017. DOI: 10.1109/TMM.2017.2759508.
- [33] K. Vikram and S. Padmavathi, "Facial parts detection using Viola Jones algorithm", in *Proc. of 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2017, pp. 1–4. DOI: 10.1109/ICACCS.2017.8014636.
- [34] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I-511–I-518. DOI: 10.1109/CVPR.2001.990517.
- [35] M. Gupta, S. Kumar, N. Kejriwal, L. Behera, and K. S. Venkatesh, "SURF-based human tracking algorithm for a human-following mobile robot", in *Proc. of 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2015, pp. 111–116. DOI: 10.1109/IPTA.2015.7367107.
- [36] S. An, X. Ma, R. Song, and Y. Li, "Face detection and recognition with SURF for human-robot interaction", in *Proc. of 2009 IEEE International Conference on Automation and Logistics*, 2009, pp. 1946–1951. DOI: 10.1109/ICAL.2009.5262624.
- [37] K. E. A. van de Sande, J. R. R. Uijlings, T. Gevers, and A. W. M. Smeulders, "Segmentation as selective search for object recognition", in *Proc. of 2011 International Conference on Computer Vision*, 2011, pp. 1879–1886. DOI: 10.1109/ICCV.2011.6126456.
- [38] R. Alimuin, A. Guiron, and E. Dadios, "Surveillance systems integration for real time object identification using weighted bounding single neural network", in *Proc. of 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017, pp. 1–6. DOI: 10.1109/HNICEM.2017.8269461.
- [39] Y. Pang, M. Sun, X. Jiang, and X. Li, "Convolution in convolution for network in network", *IEEE Trans. Neural Netw. Learn. Syst.*, no. 99, pp. 1–11, 2018. DOI: 10.1109/TNNLS.2017.2676130.
- [40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", in *Proc. of the 25th International Conference on Neural Information Processing Systems, USA*, 2012, vol. 1, pp. 1097–1105.
- [41] W. Ding, K. Liu, F. Cheng, and J. Zhang, "Learning hierarchical spatio-temporal pattern for human activity prediction", *J. Vis. Commun. Image Represent.*, vol. 35, pp. 103–111, Feb. 2016. DOI: 10.1016/j.jvcir.2015.12.006.
- [42] H. Liu, Y. Wang, W. Ji, and L. Wang, "A context-aware safety system for human-robot collaboration", *Procedia Manuf.*, vol. 17, pp. 238–245, Feb. 2018. DOI: 10.1016/j.promfg.2018.10.042.
- [43] D. Andriukaitis, A. Laucka, A. Valinevicius, M. Zilys, V. Markevicius, D. Navikas, R. Sotner, J. Petrzela, J. Jerabek, N. Herencsar, D. Klimenta, "Research of the Operator's Advisory System Based on Fuzzy Logic for Pelletizing Equipment", *Symmetry* vol. 11, iss. 11, art. no. 1396, p. 1-17, 2019. DOI: doi.org/10.3390/sym11111396.
- [44] B. Mocanu, R. Tapu, and T. Zaharia, "When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition", *Sensors*, vol. 16, no. 11, Oct. 2016. DOI: 10.3390/s16111807.
- [45] J. Han, N. Campbell, K. Jokinen, and G. Wilcock, "Investigating the use of n-verbal cues in human-robot interaction with a Nao robot", in *Proc. of 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, 2012, pp. 679–683. DOI: 10.1109/CogInfoCom.2012.6421937.
- [46] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", 2014. [Online]. Available: ?/paper/Rich-Feature-Hierarchies-for-Accurate-Object-and-Girshick-Donahue/009fba8df6bbca155d9e070a9bd8d0959bc693c2
- [47] R. Girshick, "Fast R-CNN", in *Proc. of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169. [Online]. Available: https://arxiv.org/pdf/1504.08083.pdf
- [48] Tinkerkit Braccio robot, Arduino Official Store. [Online]. Available: https://store.arduino.cc/tinkerkit-braccio-robot
- [49] "Buy a Camera Module V2", Raspberry Pi". [Online]. Available: https://www.raspberrypi.org
- [50] Gravity: URM09 Ultrasonic Sensor (I²C), (V1.0)-DFRobot. [Online]. Available: https://www.dfrobot.com/product-1832.html?search=Gravity%20URM09&description=true

- [51] Stepper mottor JK42HM48-1684 400 speps/rev 3,0 V, Sklep dla robotykow. [Online]. Available: <https://botland.com.pl/pl/silniki-krokowe/3606-silnik-krokowy-jk42hm48-1684-400-krokobr-30v-168a-043nm.html>
- [52] "Buy a Raspberry Pi 3 Model B+", Raspberry Pi". [Online]. Available: <https://www.raspberrypi.org>
- [53] Power supply Mean Well RS-15-5 - 5 V/3 A/15 W, Sklep dla robotyków. [Online]. Available: https://botland.com.pl/pl/zasilacze-montazowe/5840-zasilacz-montazowy-mean-well-rs-15-5-5v-3a-15w.html?search_query=zasilacz+%225V%2F2A%22&results=454
- [54] Workstation Dell Precision Rack 7910, Dell Polska, Dell. [Online]. Available: <https://www.dell.com/pl/firmiinstytucji/p/precision-7920-workstation/pd>
- [55] R. Sotner, O. Domansky, J. Jerabek, N. Herencsar, J. Petrzela, D. Andriukaitis, "Integer-and fractional-order integral and derivative two-port summations: practical design considerations", *Applied Sciences*, vol. 10, iss. 1, art. no. 54, pp. 1-25, 2020. DOI: 10.3390/app10010054.
- [56] Jetson TX2 Module. [Online]. Available: <https://developer.nvidia.com/embedded/jetson-tx>