A Review of Art and Real World Applications of Intelligent Perception Systems

[1]Ceren Ergenc and [2]Yifei LI
[1,2] Communication Science, University of Amsterdam, 1012 WX Amsterdam, Netherlands.
[1]ceranergenclkp@gmail.com

ArticleInfo
R. Arulmurugan et al. (eds.), *First International Conference on Machines, Computing and Management Technologies*, Advances in Intelligent Systems and Technologies
Doi: https://doi.org/10.53759/aist/978-9914-9946-0-5_9

Abstract – Sensory data and AI/ML techniques are crucial to several robotics applications, which is why perception in robots is a hot topic. Some of these applications include: object recognition, scene understanding, environment representation, activity identification, semantic location classification, object modeling, and pedestrian/human detection. Robotic perception, as used in this article, is the collection of machine learning (ML) techniques and methods that allow robots to process sensory data and form conclusions and perform actions accordingly. It is clear that recent development in the field of ML, mostly deep learning methodologies, have led to improvements in robotic perception systems, which in turn make it possible to realize applications and activities that were previously unimaginable. These recent advancements in complex robotic tasks, human-robot interaction, decision-making, and intelligent thought are in part due to the fast development and widespread usage of ML algorithms. This article provides a survey of real-world and state of the art applications of intelligent perception systems in robots.

Keywords – Robotic Perception, Perception Systems, Environment Representation, Machine Learning, Artificial Intelligence

I. INTRODUCTION

In robotics, "perception" refers to a mode, which provides data about the robot's physical environment and trains the machine to make rational decisions based on that data. This article focuses on "weak AI," also known as "standard machine learning approaches," rather than "strong AI," which has not yet been achieved in practical robotics applications. In order for a robot to execute decisions, carry out tasks, and formulate plans in real-life, it must have access to information about its surroundings. Environment change detection, human detection, pedestrian detection, vehicle detection, road detection, terrain classification, activity classification, gesture and voice recognition, three-dimensional (3D) environment representation, semantic place classification, object recognition, object tracking, and obstacle detection are all examples of subfields of robotic perception that can include autonomous robot-vehicles. Machine learning (ML) techniques, ranging from the tried-and-true to the cutting edge of deep learning, are used in the majority of today's robotic vision systems.

Robotic perception may be achieved by a variety of machine learning techniques, including supervised learning with hand-crafted features, unsupervised learning, deep-learning neural networks, and, and hybrids of these and other techniques [1]. Irrespective of the particular ML technique employed, data from the sensors is/are a crucial part of robotic perception. Data may be gathered from a number of different places, including the robot itself, surrounding objects, or even another robot (such as a camera mounted onto an unmanned aerial vehicle (UAV) nearby). However, an efficient method to aggregate and interpret data from a sensor is fundamental before a ML technique can be applied in a situation where several sensors are engaged in perception (whether they are all using the same modality or are all using distinct modalities).

Procedures for data calibration and alignment may be fundamental, dependent on the condition of the issue and the sensors being employed. The ability to create a mental model of the environment from sensor input is an essential part of a robot's perceptual system. To acquire a metric framework and its semantic representations is what we mean by "mapping," thus we'll use that term interchangeably with "environment/scene representation" in this context. Multiple stages of machine learning are used in this semantic mapping process, such as reasoning about volumetric occlusions, and accuracy, and describing, detecting and optimally correlating localized regions from various time-stamps/frameworks. Nonetheless, the major purpose of ecological mapping is to depict data collected by the robot's own exteroceptive sensors so that the robot may draw conclusions and make assessments about its immediate surroundings.

Robots engage in highly context-dependent perceptual tasks such as localization and navigation. A robot's primary function is irrelevant to whether it is inside or outside. The mapping (representation) and perceptual systems may, therefore, make different assumptions when dealing with indoor vs. an outdoor setting. Furthermore, a perception system's sensory input to be processed will change substantially whether it is running inside or outside owing to the different sensors used in each environment. The differences and challenges encountered by a mobile robot while working inside vs outside may be seen in the surface or terrain on which it operates. However, field (outside) robots have the challenge of modeling an environment that is sometimes far from regular, and without an appropriate representation, subsequent perception tasks are hampered. The environment representations models are helped by the fact that many indoor robotic systems assume that grounds are typically level and regular. In addition, the visual system of an outdoor-operating robot must adjust to variations in temperature, humidity, wind speed and direction, and light intensity and spectrum. Participants in the 2016 Amazon Picking Challenge voted robotic vision as one of the major challenges in the place-and-pick application field because to the identical scenario-specific variances found in nearly all applications of the robotic vision.

One of the 2016 teams benchmarked a posture estimation algorithm on a warehouse operations dataset and found that the technique's performance varied greatly depending on the quantity of clutter available and the kind of item being evaluated. Therefore, specialized knowledge is required for the selection, adaptation, extension, and refinement of the many components employed in modern perception systems. Since it is generally feasible to gain the necessary findings straight from raw data, by creating massive data, the end-to-end learning component of most deep-learning algorithms has facilitated the creation of perception paradigm for beginners. The framework handles the laborious process of recognizing features, characterizing them, filtering them, matching them, and optimizing them, thus selecting a method often requires downloading a newly pretrained network from a database and customizing it to meet the circumstance at hand. The necessity for large amounts of training data is self-evident given that there is currently no ready-made DL technique for each problem, or at least ineffective pre-learned network.

As a result, modern AI/ML relies heavily on having access to large datasets. Shaikh and Chai [2] give an overview of RGB-D datasets and provide methods for synthetically building sensor-based datasets that may be used for perception tasks. Overfitting to these criteria, however, is possible since the deployment ecosystem of mobile robots is not the same as the one applied to train robots to understand and perceive its surroundings. Consequently, Mastrogiovanni, Sgorbissa, and Zaccaria's [3] suggestions should be taken seriously by both academics and professionals. One major difference between traditional computer vision perception and robotic perception (identified as robotic vision in [4]) is that in robotics, the outcomes of a perception system's outputs will influence real-world choices and actions. An active, embodied, complex and goal-oriented robotic system relies heavily on perception for its success. Robotic perception, as shown by Morillo-Mendez, Schrooten, Loutfi, and Mozos [5], requires the translation of pictures (or point-clouds or scans) into actions, while many computer vision systems just record photographs and convey the results as data.

In this article, we'll take a look at the state of the art and emerging trends in intelligent perception systems for robotics, including topics like environment representation, applications of AI and ML to robotics perception, and related use-cases. Below is the outline for the rest of the article: Section II presents a discussion of environment representation and network's self-awareness. Section III focuses on machine learning and artificial intelligence on robotics perception. In Section IV, four case studies (RobDREAM project, Strands project, AUTOCITS project, and SPENCER project) have been discussed. Lastly, Section V draws final remarks to the paper.

II. ENVIRONMENT REPRESENTATION

As a type of internal representation, a network may help solve a path planning problem by locating and labeling open, obstacle-free zones on a map as possible endpoints along the route. In the map, each delayvector is represented by a node, and the node's attributes are two variables related to the input position (x, y). No changes are necessary to apply the concept to the third dimension. It is taken for granted that mobile robots have some kind of tracking mechanism, such a GPS or odometer, to keep tabs on their whereabouts. The robot should have simple features for detecting and avoiding obstacles.

Six-node processing networks are connected to create a network topology that may be thought of as an internal map and used to explain pathways and paths taken by robots. The robot has to plan its route from the starting point to the final goal, taking into account the presence of obstacles along the way (path planning algorithm is discussed in next section). At each step of mobility, the robot's position coordinates are sent into the network for simulation and fine-tuning. The network nodes will begin to move to previously unoccupied spaces where robots are present. Working off of a known map, the robot's internal representation may be developed in a manner that reduces the need for it to visit each and every location. In order to traverse uncharted regions of the globe, robots may engage in a search and exploration phase, starting and terminating at predetermined points or at sites chosen at random.

In this case, it is helpful if the starting and ending points are quite near together. Occupancy grid mapping is the most common technique used among the many approaches used to describe the environment for autonomous robotic vehicles and mobile robots. This two-dimensional mapping is still employed in most mobile systems since it is effective, has a probabilistic foundation, and can be implemented quickly. The use of 2.5D and 3D models has replaced the use of 2D photographs in many applications; nevertheless, the use of 3D models is becoming more common. There are essentially two key motives for making use of higher dimensional representations: Two causes have led to the development of 3D environment representations: (1) robots are planned to make decision and move in increasingly sophisticated circumstances whereby 2D demonstrations are ineffective, and (2) modern 3D sensor techniques are trustworthy and affordable.

In addition, the development of software libraries like as PCL and ROS, and the emergence of Author-developed approaches such as Octomaps, have all led to the prevalence of 3D-oriented environment representations. With the widespread availability of RGBD sensors, cartographers have the potential to produce maps with more detail and scope than ever before. Moreover, a lot of efforts have been made on semantic labels of maps, booth voxel and pixel levels. The bulk of the applicable methods may be grouped into two overarching categories: online and offline.

Fig 1. Network's self-awareness occuring in stages

In **Fig. 1**, each stage being triggered by a unique combination of inputs (black dots). An improved layout of nodes is depicted after several training iterations. A semantic map is built in real time from the data collected by the mobile robot. Together, this toolkit and a simultaneous localization and mapping (SLAM) framework guarantee a consistent geometric shape for the final map. The ability of robots to create reliable maps of their environments is a hotspot for study because of how important it is to their functionality. In an effort to address the SLAM challenge, researchers have combined two previously independent methods for doing so for the first time. To incorporate and deal with time dependencies (long- and short-term) into underlying structures, recent efforts have used grid maps, normal distribution transform (NDT), and pose-graph representations. Wang, Qin, Cheng, Zhu, Wang, and Zhu [6] describe how they used random forest classifiers to forecast semantic labels from RGBD data, and how they then regularized these labels using a conditional random field (CRF) method. To create a map that is both accurate and consistent geometrically, Kragh and Underwood [7] use their own elastic fusion SLAM method to combine CNN predictions about the scene. Jiang et al.'s [8] work makes use of convolutional neural networks (CNNs) to progressively construct a semantic map, and they plan to enhance the CNN's class support by incorporating a set of online-trainable one-vs.-all classifiers.

There are a number of semantic mapping techniques available, all of which can take a global map as input and work locally, even in the absence of an internet connection. By following Simanjuntak and Simanjuntak's [9] procedures, the input data is partitioned, and the resulting "rooms" are displayed. Mura, Mattausch, Villanueva, Gobbetti, and Pajarola [10] computed the segmentation using a 2D cell-complex graph-cut approach, but this method is limited to one-story buildings; Wang, Yang, Shen, Ma, Zheng, and Fan [11] process multi-story buildings by identifying the gaps between floors, walls, and ceilings, but this method requires that the construction walls be aligned along a single axis. The technique presented by Simão, Gomes, Alves, and Brito [12] utilizes 3D cell-complex models to assess a larger-point cloud of indoor structures, with the output mesh integrating the semantic segmentations of input dataset. The major problem is that is needs prior knowledge of where the scans were performed in order to obtain the input data.

Semantic fragmentation of the ecosystem has recently been described by Bellos, Basham, Pridmore, and French [13], who extend on previous work by evaluating techniques to integrate multiple forms of data, such as the availability of goals and indications of distinct room types. The work by DasGupta and Shaw [14] sought to achieve a human-life and understandable classification of the environment while keeping as numerous of the semantic aspects as feasible. Petrović, Nikolić, Jovanović, and Delibašić [15] also resort to inferring from a conditional random field (CRF) or fusing several different types of data using the Gibbs sampling technique. Improving robots' perception and representation of their environments has been a central focus of robotics research, especially for automated driving models (or autonomous robotic vehicles). Metric representations (3D or 2D) to abstract topology maps are discussed, as are other methods for manipulation, object recognition, navigation, localization, etc.

The capacity to provide a mathematically accurate map that is annotated with semantic data has applications outside the realms of building management and design; such a map might be sent back into the robotic model to enhance the latter's situational recognition and, by extension, its capability to undertake particular tasks in a human-based environment (e.g., If the robot already has an idea of the location of the kitchen, it will have a better chance of locating a cup).

III. MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE ON ROBOTICS PERCEPTION

Artificial Intelligence in Robotics

Because of its versatility and capacity for learning, artificial intelligence (AI) is quickly becoming the preferred component in many robotic systems. Intelligent machines are now within grasp. Incredibly increased process adaptability and flexibility is made possible by artificial intelligence. AI-enabled robots serve as a link, if you will, between the robotics and AI communities. Robots operate under the control of an artificial intelligence (AI). Artificial intelligence (AI) algorithms are used in robotics programming to enable the robot to do increasingly complex tasks. A path-finding approach may be used by a warehouse robot to go about the warehouse. A question that has been raised is whether or not AI systems have access to all the data required for different types of reasoning. Robotics, however, takes AI out of the digital and into the real world, where it may engage in real-time interaction with the actual environment around it.

Robots are becoming more lifelike as researchers use machine learning and artificial intelligence to enhance their sensory capabilities. Artificial intelligence (AI) and robotics engineering have very little in common with one another. Robotics, however, include not only the construction and operation of physical robots, but also the conceptualization, development, and use of virtual robots. Both of the most popular AI frameworks can work with automated systems. Software (through microprocessors and microcontrollers) provides the first kind of intelligence, instructing the hardware on what to do and how to make decisions. The more it's used, the more the program improves itself. Hardware intelligence is the other kind of intelligence robots possess; it enables them to mimic human mental processes by way of learning circuits.

Innovations in mechatronics, electrical engineering, and computing are enabling roboticists to create robots with more sophisticated sensorimotor capabilities that can adapt to their changing environments. The machine has always been at the heart of the adaptive and highly precise industrial production system, which tolerates very little in the form of variation. There has never been a more convenient time to implement in established systems. Environmental autonomy is comprised of a robot's ability to see its surroundings, make plans on how to respond, and then carry out those plans (manipulating, navigating, and collaborating). The fundamental motivation for combining AI with robotics is to give robots more freedom of movement by giving them access to AI's vast store of knowledge. One sign of this sort of intelligence is the capacity for foresight in the context of task planning or interaction (manipulation or navigation) with the external environment. There have been several attempts to build sentient robots. There are now robots that can drive cars, fly in both natural and man-made settings, swim, move boxes and supplies over a range of terrains, and pick up and put items, but building a system with human-level intelligence is still years away.

Perception is one of the many significant applications in AI in robotics. The perception of robots is assisted by either computer vision, or on-board sensors. Over the past few years, developments in computing have amounted to much enhanced vision and sensing. The ideology of perception is not only fundamental for planning, but is also assists in establishing a false impression of self-awareness in robotics (see **Fig. 1**). So, the robot can communicate with other objects in the vicinity. It is a branch of research known as social robotics. Among the many topics it covers are cognitive robotics and Human-Computer Interaction (HCI). HCI seeks to improve robots' ability to read and respond to human cues, such as emotions and body language, in order to facilitate more natural interactions between humans and machines. The cognitive robotics field concentrates on providing machines the capability for knowledge acquisition and autonomous learning through sophisticated perceptual mechanisms like observation, imitation, and experience. The target is to model the inner workings of the human brain in order to speed up and improve the learning and memory processes. Cognitive robotics models exist that harness the power of innate curiosity and desire to learn more quickly and thoroughly. Since then, AI has broken every record and overcome several challenges that were unimaginable only a decade ago. As a consequence of this synthesis of progress in many disciplines, our knowledge of robotic intelligence will grow and change.

Seasons/AI and Robotics

There have been a number of "springs," or periods of optimism, followed by "winter," or periods of pessimism, about artificial intelligence in Table 1.

AI Technologies and Disciplines

Each of the following is essential to the development of AI since the field spans so many disciplines. There are several probabilistic and heuristic methods to computing, such as fuzzy logic, neural networks, evolutionary computing, and more. Artificial neural networks are a branch of connectionism that attempts to recreate the brain's complex information processing mechanisms. Use of Artificial Neural Networks (ANNs) and its derivatives has led to significant progress in AI's capacity to perform "perception-related" tasks. Modern multicore parallel computing hardware platforms make it possible to layer several neural networks, each of which may learn its own set of characteristics independently of human intervention or specialized training material. Deep learning is the term for this strategy. While deep layered ANN applications have the potential to achieve significant levels of efficiency, they are sometimes hampered by 1) the lack of interpretability of the resulting learned model and 2) the need for substantial computer resources to analyze large volumes of training data.

The subject of machine learning known as "deep neural networks" is well-known for its ability to learn complex knowledge or data representations in small, incremental steps. Messages are sent from higher to lower levels of organization through these intermediary stages. These levels give more granular representations of data to aid in tracking and spotting. Various domains, including automated voice recognition, computer vision, and audio/music signal identification, have found success with deep learning systems like deep neural networks, deep convolutional neural networks, and deep belief systems. Fuzzy logic is used to manipulate data that consists of fuzzier information. For the most part, computational intelligence models account for the reality that, in many real-world situations, our knowledge of the environment is incomplete or incorrect, despite the fact that our observations are always spot-on.

Fuzzy logic is useful because it allows humans to handle data while assuming particular amounts of insightful data throughout different sets of observations, and its framework contains features that boost the model's interpretability (Zadeh, 1996). On the other hand, it provides a framework for formalizing AI approaches and a straightforward mechanism for converting AI systems into electrical circuits. Since fuzzy logic does not generate learning skills, it is often combined with additional components like evolutionary computing, statistical learning or neural networks.

Table 1. Illustrates the different springs, and periods of AI and robotics.

Table 1: Springs/periods of AI/Robotics	
Early Computer Programs (1952-1956)	Prior to the coinage of the term "artificial intelligence," curiosity in cybernetics and neural networks was already on the rise. The Dartmouth Conference (1956), a zenith of this developing interest, ushered in several years of unparalleled advancement in artificial intelligence.
Original Spring (1956-1974)	The computer systems of the day have the capacity to solve problems such as geometry, algebra, and English conversation. Experts were optimistic about the progress that had been achieved, calling it "wonderful." Researchers in this subject predict that intelligent robots will be developed within the next two decades.
First Winter (1974-1980)	The winter began when the media and the public questioned AI's promises. The scientific community was caught in a whirlpool of hope, but the limitations of the available technology were impassable. Withdrawal of support from major financial sources including DARPA, the British government and the National Research Council led to the first "winter" in the history of artificial intelligence.
Second Spring (1980-1987)	Expert models were designed to solve issues using logical concepts gleaned from specialists in a certain field. Similarly, connectionism and neural networks saw a rebirth in popularity for their potential applications in fields like pattern recognition and language processing. The second "spring" of artificial intelligence occurred about now.
Second Winter (1987–1993)	General-Purpose Desktop Computers Replace Expert System Workstations. As a result, several companies that developed expert systems went out of business. Consequently, there was a new wave of pessimism, and the financial plans that had been made in the spring were abandoned.
A brief history (1997-2000)	From that year until the year 2000, no high-profile, multi-million dollar projects were undertaken in the field of artificial intelligence. More processing power and resources were made available, and progress was achieved despite a lack of major funding. Machine learning became an essential AI principle, and new applications were developed for niche markets.
Third Spring (2000-Now)	The spread of the Internet and websites has facilitated the development of new areas like Deep Learning and Big Data. We seem to be living in what has been called the "third spring of AI." Some have projected that over the next several decades, a Singularity would occur, marking the birth of a vast super-intelligence that will one day exceeds human cognition. The question is if this is even possible.

Evolutionary computing is oriented on the key metrics of natural selection, or on earlier observed patterns regarding the behaviors of a particular classification. The two most fundamental research fields in this case are swarm intelligence, and genetic algorithms. Since its strongest suit is multi-objective optimization, it has the most impact on that particular sector of AI. These models suffer from the same interpretability and computing capacity limitations as neural networks. The discipline of statistical learning takes a more conventional statistical stance, like Bayesian modeling, by bringing the concept of prior knowledge into AI. These methods build on the basis of traditional statistical techniques and operations to provide formal approaches to AI. One important difficulty is that the probability concept is not always applicable, especially when dealing with situations where uncertainty or subjectivity must be measured. The outcome of using probabilistic methods is referred to as a "correspondence to a population" in the field.

The goal of meta-algorithms and ensemble learning, an AI branch, is to develop models by merging different, relatively weaker base learners to enhance accuracy and decrease variance and bias. For example, ensembles may provide greater wiggle room than single-model approaches for describing certain kinds of complex patterns. Two common meta-algorithms for developing ensembles are boosting and bagging. Although improved accuracy is not always guaranteed, ensembles have the potential to boost the precision of the pattern search by making use of massive computer resources to train a large number of base classifiers. The use of logic-based AI is common in tasks involving the representation and inference of knowledge in artificial intelligence. Structures known as logic programs may be created in formal logic to convey the predicate facts, descriptions, and domain semantics. From the corpus of known information, a hypothesis may be derived using inductive logic programming.

Machine Learning in Robotics

Despite efforts to mitigate the challenges of soft sensors and actuators, which cannot attain accurate controls and calibrations, there are still a number of constraints on the applications of algorithms in machine learning.

First, machine learning approaches are data-driven strategies that often need a large dataset in order to train their networks adequately. Costing a lot of time and energy to collect, large data sets may be quite useful. The reliability of the findings is further diminished if the collected data are biased or inaccurate (i.e., the data or information does not reflect the complete behavior of a robot, just portions of it). There are a variety of ways to go about trying to resolve the problem. To begin, simulations enable the gathering of massive amounts of data in a number of contexts. The use of simulation settings helps reduce the need for trial and error, which may lead to problems like robot damage. Artificial worlds have been created where soft robots can function. Nevertheless, it is still debatable whether or not these simulated environments aid in reducing the quantity of real-world training data. Non-linear autonomous robots frequently contain a large number of DOF, thus it is important to double-check simulation parameters before releasing them into the wild.

Another drawback is that the mechanical or mathematical frameworks utilized in simulated ecosystems do not always capture the nuances of how soft sensors and actuators behave in the actual world. There have been a number of publications proposing methods for making the transition from virtual to physical environments; it is essential to put them to the test in a soft robot environment. The use of machine learning techniques is another viable approach for decreasing data size. For instance, meta-learning strategies have been proposed as a way to quickly acquire knowledge with little new information. The goal of transfer learning is to speed up the learning process of a single dataset by transferring earlier acquired data from another collection of data in a similar domain. One human demonstration or video series forecasts may teach a robot arm how to operate and what policies to follow.

These techniques might be used to instruct robots equipped with soft sensors and actuators. In addition, you may use similar techniques to calibrate a second sensor/actuator with a little amount of new data if datasets for the first sensor/actuator already exist. Using hysteresis features to build a kernel function, which assesses the similarity, between the target data and the source, is one methodology of calibrating soft sensors by the use of few-short training. Soft robotics is a promising area where these methods may be used, however they have not yet been tried. Since existing studies on stiff robots using Meta learning are largely focused on vision data, it may also need a novel issue description, which is fundamental for soft sensors and actuators.

Second, even though recent researches have concentrate on problems such as hysteresis and non-linearity, there are many more types of errors, which have negative implications on the performance of soft robotics. The majority of soft sensors and actuators are developed using manual processes, which may lead to manufacturing faults inside the devices. This can have an impact on the efficiency of machine learning methods. While machine learning may be used to provide a description of a sensor or actuator, it is not known whether or not that model can be applied to other sensors or actuators without significant modification. On top of that, after some use, most soft materials will have a little deformation, which might reduce the accuracy of machine learning models. Since a machine learning technique may need to be re-trained everytime an endpoint is changed, this reduces the approach's generalizability.

This may be avoided by transferring previously taught settings to new or previously owned devices to expedite the re-training process. Three, more study is required to understand the feasibility and limitations of genuine robots in the real world. Deep learning methods have been heavily used in recent studies; however these studies need extensive calculations that can only be operated using graphics processing units (GPUs). Because of this, the robot's central processing unit will have to grow in size. As soft robots are typically small enough to be carried or worn, they cannot be made much bigger. In addition, small embedded systems cannot provide the kind of real-time, speedy calculations that are necessary in emergency situations. It becomes a much more difficult task to govern robots. Recent advances in artificial intelligence have focused on refining machine learning models to execute calculations faster without losing accuracy, suggesting that this limitation may soon be addressed.

Robotics Perception

A robot can go to work after it has (self) localized its position. For this purpose, autonomous mobile manipulators must explore their surroundings and zero in on and grasp specific targets. To produce a 3D map used for collision-free grasp categorization and item localisation, the robot typically travels to the target area, conducts a survey, and processes the collected data. Anything from a single item to a whole tabletop or container might need to be picked up and placed somewhere else. In the latter scenario, a detailed 6-DOF estimate is necessary. The next step is to plan and carry out a grab. Visual servoing, for instance, is used to synchronize perceptual and actuational processes in order to perform very accurate manipulation. Each application, however, may benefit from a more comprehensive approach that deals with perception and manipulation simultaneously. The perceptual and manipulation processes of the brain are intertwined and crucial to a complete understanding and interaction with the external environment, as Zhang, Gao, Holmes, Mavrikis, and Ma [16] of the common coding hypothesis have shown. Khan and Cañamero [17], seeing the importance of providing artificial agents with a smooth transition between perception and action, argued that computers should be equipped with "good sense organs, which money can purchase" and permit to learn from their respective failures until they overcame the challenge.

Taking into account the work of Averta, Della Santina, Ficuciello, Roa, and Bianchi [18], as well as the most recent findings presented in [17], we find that robot perception involves both planning and interactive segmentation. Particularly, identifying things through a two-way flow of information between perception and action may provide the greatest outcomes. Manipulating objects requires considering not just their segmentation but also their orientation and location relative to robots, which is why localization is an integral part of the manipulation issue. As mentioned by Presenti, Liang, Pereira, Sijbers, and De Beenhouwer [19], the issue of object pose estimation is often addressed by using precomputed grip points. Some people advocate for focusing on particular features and bug fixes, while others advocate for working with existent models and vocabularies. Conjectures derived from data are checked by concept-driven and top-down models. Saeidi and Arabsorkhi [20] investigate the widespread belief that these systems are analogous to the human visual system.

Techniques such as those described in ([21]) that employ color gradients, color histograms, normal or depth orientations from discrete item perception are all examples of camera/vision-based robotic perception. Obstructions, the impact of aspect ratio, and the complications of discretizing the 6D or 3D search spaces are all common issues for vision-based perception systems. Predictions of the object's position are instead generated using a PnP algorithm or through voting. The performance usually drops if the item being inspected has no texture and the backdrop is very busy. The aforementioned research makes use of deep-learning (such as CNN) and traditional ML-based learning methods (such as CNN). The significance of mobile perception and manipulation has been reviewed at a recent main computer vision conferences linked with SIXD Events and Challenges such as the Amazon Robotics Challenge.

However, the available choices are either too laborious or error-prone for usage in industrial settings, or they are so specific to a particular use case that their implementation requires special engineering. To effectively apply laboratory-learned models to the actual world and the unknown (novel) surroundings, transfer learning (in the sense of generalization growth) is essential. Improved accuracy (in the form of better classification or recognition performance) and faster processing times are both within the reach of deep learning. Not just in sensor technologies, as well as in LiDAR-oriented perceptual circumstances, domain randomized and domain adaptability (including image supplementation) seem to be an important avenue to examine and develop. Robots that are mobile and capable of manipulation often use their own navigation and manipulation skills to learn more about the items they are manipulating. To improve the object model estimation, one may model an item by physically manipulating it, or one can examine it from a variety of angles. To compensate for their limited sensory systems, mobile robots and autonomous (robotic) vehicles must undergo extensive offline training. Environment representation (including multisensory fusion) is very important for autonomous driving applications because to the complexity of the problems it must solve.

The topic of improved perception for (fully) automated driving has lately witnessed a renaissance in attention from the automotive and robotics sectors as well as academic institutions, thanks in large part to the European ELROB challenge (since 2006) and DARPA Challenges (2004 to 2007). Numerous studies on autonomous robot-cars, or self-driving automobiles, have been presented at prestigious robotics conferences and published in top-tier robotics journals. The foundations of V2X-based communication technologies, such as autonomous driving systems (ADS), incorporate environmental representation/modeling, and sensor fusion), localization (position identification), navigation (path control, planning, trajectory following), and, more lately, collaborations (V2X-oriented communications). ADS perception models depends majorly on the “segmentation,” recognition/detection of objects such as lane-markings, road, road, pedestrian and other road users (e.g., bicycles), other cars, traffic lights, crosswalks and the dissimilar other categories of items, and barriers present on the roadways.

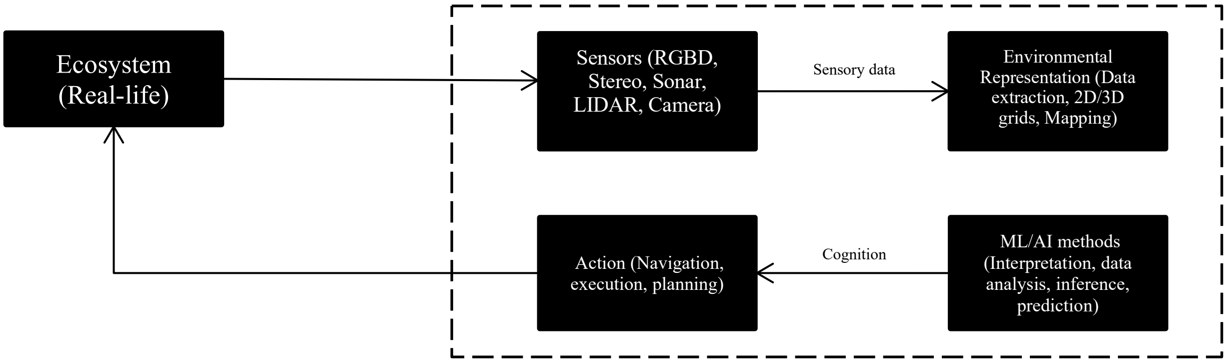
Ecosystem (Real-life)
Sensors (RGBD, Stereo, Sonar, LIDAR, Camera)
Sensory data
Environmental Representation (Data extraction, 2D/3D grids, Mapping)
Action (Navigation, execution, planning)
Cognition
ML/AI methods (Interpretation, data analysis, inference, prediction)

Fig 2. Essential parts of any robotic perception system as the processing of sensory data (in this case, primarily visual and range perceived notion), the construction of task-appropriate data models, the development of algorithms for analyzing and interpreting that data (using ML/AI techniques), and the performance of different actions required for robots to interact with its environment.

Feature extraction and object identification in ADS rely heavily on machine learning, which is used with a wide variety of sensors (including LIDAR, camera systems, Radar, "new" solid LiDAR techniques). An intriguing new trend in robotics and autonomous vehicles in the integration of cooperate data from the inter-linked infrastructure and the environment into the robotic vision system (see **Fig. 2**). By enhancing the perceptual apparatus, we want to raise dependability and security. By way of illustration, an autonomous vehicle might gain by having information about an item or obstacle on the road given to it in advance (for instance, through a communication system), just seconds before the item or obstacle reaches the field of view of the car's onboard sensors.

IV. CASE STUDIES

The RobDREAM project

For sophisticated robots to safely traverse dynamic surroundings, assess scenes, recognize important things, and control them without clashing with anything, intelligent perception algorithms are required. Currently, mobile manipulation systems' perception usually fails because of differences in the context (such as the illumination, the items being utilized, the manipulation surroundings, or the areas). Then, robotics experts must either choose a different approach or sensor, or modify the settings of the perception approach and the sensors being utilized. Therefore, a high degree of cognitive capacity, such as the ability to reflect on previous acts and modify course, is required to function side by side with humans. This adaptability in the face of new challenges necessitates a number of machine learning methods, such as memory for lifelong learning, annotated data for supervised learning through users' engagements, Bayesian optimization to remove brute-force searches in different high-dimensional data and representation of meta-data for streamlined expertise sharing. RobDREAM is a group that standardized and automated a lot of these processes. To show how automatically modifying task application pipelines as per user-defined performance approach paves the way for straightforward programming and simpler deployment of robotic applications, the H2020 RobDREAM project, financed by the European Union, uses a mobile manipulator.

These annotations are employed by a Bayesian optimization models to fine-tune the pipeline of stock for each new situation the robot meets, thereby raising the bar for the system's performance. Perception was one of the important mobile manipulation technologies investigated for this research. Different strategies, including Bayesian optimization, were used to fine-tune the robot's navigation, manipulation, and grasping skills while its perceptual abilities remained unchanged. Nonetheless, the combinatorial complexities of interlinked space parameters of procedures integrated proved challenging for even the most proficient meta-learners. Two RBD-D cameras were used to generate a publicly accessible pose-annotated database that showcases a demo with practical industrial use in the installation and kitting of electrical cabinet parts.

The Strands project

The EUFP7 Strands project represents an interconnection between two private companies and six universities. The key objective of projects is to develop complex mobile robots, which are capable of operating alongside humans for a longer timeframe without being in danger. The field of mobile robotics has seen significant development over the last several decades, but robots that can perform consistently and for increased duration of time in human-centric ecosystem are still fundamental. Strands purpose to fulfill this need by providing intelligent, durable robots that can assist with a wide variety of useful jobs, from home security to senior care. Given the extended length of operations, it is crucial that the created robotic systems be able to handle with ever-changing big datasets, and the unstructured and complex real-life.

Fig. 3 depicts the fundamental functioning of the Strands system, in which the mobile robot shifts independently between waypoint series. A task organization system determines where and when the robots have to travel each day based on the jobs it has to do. The perceptual system is fundamentally a module that generates local metric maps at the robot's endpoints. These local maps are not only utilized to classify things as either mobile or stationary, but they are also continuously refined when the robot revisits the same locations. As an example of a high-level activity that is prompted by dynamic segmentations, the robot may move around a detected item to gather additional data that is then incorporate into a canonical object representation. The data is therefore employed to train a convolutional neural network, which may be used to reliably identify the item in future observations.

Spectrum analysis (that is, executing a Fourier transform onto raw data for recognition) as defined in [22] may be used to take use of the observed dynamics in the environment in order to spot patterns, as can a multitarget tracking network constructed on Rao-Blackwellized particle filters. Strands are a perceptual framework that can identify and represent not just physical objects and environments, but also people. Truong, Yoong, and Ngo [23] combine RGB-D and laser to effectively identify humans and permit human-aware navigation methodologies, both of which make robots friendly, while Ma and Wang [24] propose a methodology to continuously approximate the head-pose of individuals. Biradar, Shiparamatti, and Patil [25] provide convolutional neural network (CNN) based system for object recognition by use of laser scanner datasets; the example scenario shows the approach's usefulness by locating mobility aids like walkers and wheelchairs. One of the core features of the Strands system is the implementation of reliable long-term perception algorithms.

Since the robot will make more observations and collect more data as it explores its surroundings, any method it employs must be both trustworthy and extensible. Making such a robotic model function would require a perception stack, which is capable to progressively integrate real-life data, extract essential components, and construct models that grasp and can predict the environment. Understanding both place and time is crucial for mobile robots because it allows them to distill the data they collect during days of autonomous operation into frameworks, which can be employed to develop their functionalities over time. Cronie and Mateu [26] developed spatio-temporal frameworks of the ecosystem and utilize them for development using a data-theoretic methodology that foretells the possible accomplishment of perceiving certain parts of the globe at various times, whereas Jabeur, Ballouk, Arfi, and Khalfaoui [27] of the former paper focus on modeling environmental periodicities and assimilating them into a scheduling pipeline.

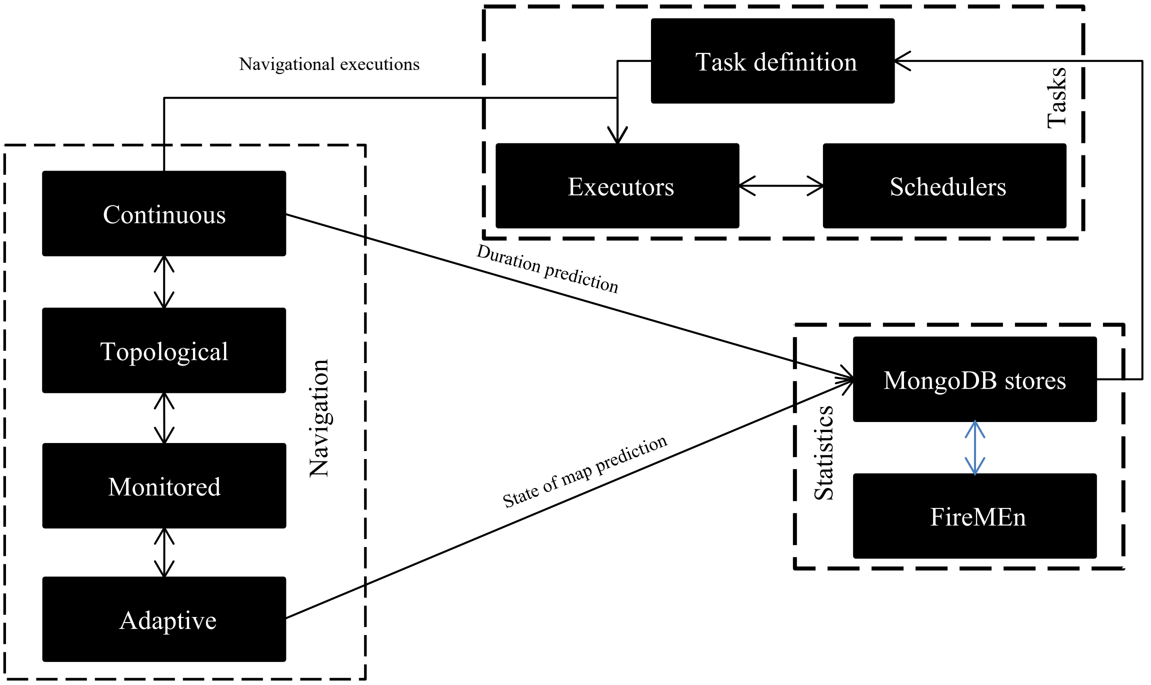

Fig 3. An overview of the Strands system

The AUTOCITS project

The AUTOCITS project will use real-world Pilots to explore and assess regulations relevant to cooperative systems and autonomous driving, as well as undertake in-depth evaluations of these technologies. With support from the European Union's CEF (Connecting Europe Facility) project, the AUTOCITS project is developing Cooperative/connected ITS (intelligent transport systems) to facilitate two-way data exchange between vehicles and infrastructure through V2V and V2I communication technologies, thereby easing the application of autonomous vehicles on smart roads. The "Atlantic Corridor of the European Network," which encompasses the major European cities of Madrid (Spain), Paris (France), Lisbon (Portugal), will be employed for testing autonomous and connected vehicles (such as low-speed robotic vehicles or autonomous shuttles). The AUTOCITS system combines several different technologies, including those used for autonomous driving both on and off the road, as well as those used in vehicles themselves (OBU, RSU).

The technologies now in use on our roads correspond to Levels 3 and 4 of automation (according to the SAE's scale for determining the degree to which automobiles are automated). The Pilot's rollout will be on the third or fourth most demanding level while utilizing AUTOCITS. A long way remains until we see completely autonomous Level 5 cars (in which the steering wheels are unnecessary) on public roads and highways. We may thus conclude that the perception system takes control of all aspects of perceiving and responding to the world around us. As a result, the autonomous vehicle's sensing, understanding, and reasoning abilities are contingent on the efficacy of the perception system, which comprises its software modules. To function in a cooperative and linked environment, V2X-enabled vehicles would not only depend on information gathered by their own in-vehicle sensors but also on information gathered by other V2X-enabled vehicles, infrastructure, and road users (and vice-versa).

The SPENCER project

It is becoming more important for robots to have "socially aware" features in contexts where they will be working alongside a large number of people. This person respects the personal space (and privacy) of others with whom they come into touch by not using set navigation to divide cues or groupings, etc. Most robotics laboratories don't have the resources to develop such features, and institutions that focus on user experience often don't have the funds to develop really innovative robots. But the European Union's Seventh Framework Programme (FP7) supported SPENCER, a multidisciplinary effort led by end-users within the aviation sector.

Since more than 80% of all passengers at major airports are in transit between flights, KLM has an interest in the efficient management of passenger movement at hubs such as Amsterdam's Schiphol Airport. For example, passengers may miss their flights if transfer times are short and the airport is huge and unfamiliar, or if they have trouble communicating with airport staff due to language or alphabet barriers. In these cases, robots that can be swiftly deployed and reserved might be useful. There is a demand for solutions like the SPENCER prototype's smart passenger flow organization and mobile data provider, which looked into this area of application (see **Fig. 4**).

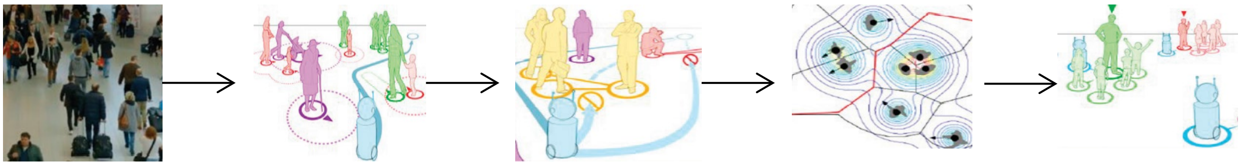

Fig 4. SPENCER project concepts and results

The SPENCER partnership has incorporated the found technologies onto a robotic system that picks up groups of passengers with short transfer times at their boarding gate, recognizes them with inbuilt boarding pass scanners, takes them to Schengen barriers, and asks them to utilize priority tracks. In addition, a KLM information kiosk is available for passengers' use. Short and secure pathways for mobile robots might be hard to construct in busy environments like airports. Therefore, the interpretation of social settings and predictions of human mobility in crowds are not fully addressed, while being crucially vital for any robots, which have to efficiently navigate in human contexts, perhaps under time restrictions. Accurate monitoring and forecasting of individuals' movements in a social situation may be challenging if there are many impediments to movement or unexpected changes in the way people are moving. Traditional route planning algorithms produce a robot that is either too restricted or too cautious to construct a safe and viable path among the throng, or that organizes a sub-optimal and enormous detour to remove people from the scene.

V. CONCLUSIONS

Current advances in robotics and artificial intelligence are limited to particular application. One of the limitations of AI is that it cannot "use common sense," or make decisions based on information that falls beyond the parameters of its training. One contemporary example is Microsoft's Toy, an AI robot developed lately for use in online discussions. Soon after its debut, it was pulled down because of its inability to differentiate between negative and positive human interactions. Also lacking is emotional intelligence; a field in which artificial intelligence has struggled to date. The only human emotions that can be identified by AI are neutrality, tension, pain, fear, sadness, joy, and anger. Emotional sensitivity defines one of the most recent frontiers of self-actualization. Authentic artificial intelligence does not exist at this time. For AI to reach this level of intelligence, it will need to mimic human cognition by learning to think, dream, feel emotions, and have independent objectives. Even though there is minimal evidence to indicate that complete AI will exist before 2050, it is fundamental to consider the effective of AI not just from a technological standpoint, but also from an ethical, legal and social standpoint.

Comparatively, current convolutional neural networks (CNNs) attain super-human categorization performance on selected domains, whereas traditional vision could only reach the performance level of a child (e.g., ImageNet Large-Scale Visual Recognition). There has been a recent uptick in the popularity of deep-learning strategies for perceiving the world around us, which has led to significant performance improvements on different tasks such as semantic segmentation, object recognition, and identification, etc. Offline testing on publicly-accessivle data and comparisons of various approaches through typical benchmarks and contests make these advancements achievable when working on perception systems.

Deep learning (DL) has become a major over-utilized phrase in robotic conferences held in the past few years, and it has widespread support from the robotics community. Despite the fact that the filters used by CNNs may be viewed as traditional to the operations of the virtualized cortex and therefore be understood as Gabor filters, DL is currently a purely non-symbolic approach to ML/AI and is not intended to produce complicated ML/AI. The topic of autonomous driving, which links the fields of robotics with computer vision, provides a particularly compelling early example of its use. With the advent of new robotics-related technologies, formerly difficult tasks may now be automated, including response systems and visual question, activity identification and video captioning, and large-scale tracking in movies and human recognition.

References

[1]. S. E. Navarro et al., "Proximity perception in human-centered robotics: A survey on sensing systems and applications," IEEE Trans. Robot., vol. 38, no. 3, pp. 1599–1620, 2022.

[2]. M. B. Shaikh and D. Chai, "RGB-D data-based action recognition: A review," Sensors (Basel), vol. 21, no. 12, p. 4246, 2021.

[3]. F. Mastrogiovanni, A. Sgorbissa, and R. Zaccaria, "Extending the capabilities of mobile robots through knowledge ecosystems," in 2007 International Symposium on Computational Intelligence in Robotics and Automation, 2007.

[4]. M. S. Qureshi, P. Singh, and P. Swarnkar, "Intelligent fuzzy logic-based sliding mode control methodologies for pick and drop operation of robotic manipulator," Int. J. Comput. Vis. Robot., vol. 12, no. 5, p. 549, 2022.

[5]. L. Morillo-Mendez, M. G. S. Schrooten, A. Loutfi, and O. M. Mozos, "Age-related differences in the perception of robotic referential gaze in human-robot interaction," Int. J. Soc. Robot., pp. 1–13, 2022.

[6]. Y.-W. Wang, C.-Z. Qin, W.-M. Cheng, A.-X. Zhu, Y.-J. Wang, and L.-J. Zhu, "Automatic crater detection by training random forest classifiers with legacy crater map and spatial structural information derived from digital terrain analysis," Ann. Am. Assoc. Geogr., vol. 112, no. 5, pp. 1328–1349, 2022.

[7]. M. Kragh and J. Underwood, "Multimodal obstacle detection in unstructured environments with conditional random fields: KRAGH and UNDERWOOD," J. Field Robot., vol. 37, no. 1, pp. 53–72, 2020.

[8]. X. Jiang et al., “Characterizing functional brain networks via Spatio-Temporal Attention 4D Convolutional Neural Networks (STA-4DCNNs),” Neural Netw., vol. 158, pp. 99–110, 2023.

[9]. O. V. Simanjuntak and D. C. Simanjuntak, “Studentsâ€TM vocabulary knowledge: Comparative study enhancing between Semantic Mapping and Diglot Weave Techniques,” acuity, vol. 3, no. 2, p. 12, 2018.

[10]. C. Mura, O. Mattausch, A. J. Villanueva, E. Gobbetti, and R. Pajarola, “Robust reconstruction of interior building structures with multiple rooms under clutter and occlusions,” in 2013 International Conference on Computer-Aided Design and Computer Graphics, 2013.

[11]. H. Wang, F. Yang, B. Shen, K.-J. Ma, T.-H. Zheng, and Y.-H. Fan, “Construction process analysis for a multi-story building structure with floors slab of long-span,” Staveb. Obz. - Civ. Eng. J., vol. 28, no. 3, pp. 404–419, 2019.

[12]. D. Simão, C. M. Gomes, P. M. Alves, and C. Brito, “Capturing the third dimension in drug discovery: Spatially-resolved tools for interrogation of complex 3D cell models,” Biotechnol. Adv., vol. 55, no. 107883, p. 107883, 2022.

[13]. D. Bellos, M. Basham, T. Pridmore, and A. P. French, “Temporal refinement of 3D CNN semantic segmentations on 4D time-series of undersampled tomograms using hidden Markov models,” Sci. Rep., vol. 11, no. 1, p. 23279, 2021.

[14]. R. DasGupta and R. Shaw, “Cumulative impacts of human interventions and climate change on mangrove ecosystems of South and southeast Asia: An overview,” J. Ecosyst., vol. 2013, pp. 1–15, 2013.

[15]. A. Petrović, M. Nikolić, M. Jovanović, and B. Delibašić, “Gaussian conditional random fields for classification,” Expert Syst. Appl., vol. 212, no. 118728, p. 118728, 2023.

[16]. J. Zhang, M. Gao, W. Holmes, M. Mavrikis, and N. Ma, “Interaction patterns in exploratory learning environments for mathematics: a sequential analysis of feedback and external representations in Chinese schools,” Interact. Learn. Environ., vol. 29, no. 7, pp. 1211–1228, 2021.

[17]. I. Khan and L. Cañamero, “The long-term efficacy of ‘social buffering’ in artificial social agents: Contextual affective perception matters,” Front. Robot. AI, vol. 9, p. 699573, 2022.

[18]. G. Averta, C. Della Santina, F. Ficuciello, M. A. Roa, and M. Bianchi, “Editorial: On the planning, control, and perception of soft robotic end-effectors,” Front. Robot. AI, vol. 8, p. 795863, 2021.

[19]. A. Presenti, Z. Liang, L. F. A. Pereira, J. Sijbers, and J. De Beenhouwer, “Fast and accurate pose estimation of additive manufactured objects from few X-ray projections,” Expert Syst. Appl., vol. 213, no. 118866, p. 118866, 2023.

[20]. M. Saeidi and A. Arabsorkhi, “A novel backbone architecture for pedestrian detection based on the human visual system,” Vis. Comput., vol. 38, no. 6, pp. 2223–2237, 2022.

[21]. T. Peynot, S. Monteiro, A. Kelly, and M. Devy, “Editorial: Special issue on alternative sensing techniques for robot perception,” J. Field Robot., vol. 32, no. 1, pp. 1–2, 2015.

[22]. D. L. Tkachev, “Spectrum and linear Lyapunov instability of a resting state for flows of an incompressible polymeric fluid,” J. Math. Anal. Appl., vol. 522, no. 1, p. 126914, 2023.

[23]. X.-T. Truong, V. N. Yoong, and T.-D. Ngo, “RGB-D and laser data fusion-based human detection and tracking for socially aware robot navigation framework,” in 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2015.

[24]. B. Ma and T. Wang, “Head pose estimation using sparse representation,” in 2010 Second International Conference on Computer Engineering and Applications, 2010.

[25]. M. S. Biradar, B. G. Shiparamatti, and P. M. Patil, “Fabric defect detection using deep convolutional neural network,” Opt. Mem. Neural Netw., vol. 30, no. 3, pp. 250–256, 2021.

[26]. O. Cronie and J. Mateu, “Spatio-temporal c\\`adl\\`ag functional marked point processes: Unifying spatio-temporal frameworks,” arXiv [math.ST], 2014.

[27]. S. B. Jabeur, H. Ballouk, W. B. Arfi, and R. Khalfaoui, “Machine learning-based modeling of the environmental degradation, institutional quality, and economic growth,” Environ. Model. Assess., vol. 27, no. 6, pp. 953–966, 2022.