Pilot Studies on Avrora Unior Car-Like Robot Control Using Gestures

Nikita Nikiforov, Tatyana Tsoy, Ramil Safin, Yang Bai, Mikhail Svinin, and Evgeni Magid

Abstract Gesture recognition is not only an important communication channel in human-human interaction but it also allows a human to communicate with other intelligent devices. This paper presents a concept for controlling the car-like robot Avrora Unior locomotion using gestures. We created a list of 18 control commands that contains basic and compound commands. A group of 17 volunteers used this list to create individual control gestures independently. A small part of the obtained dataset of gestures was used with the Teachable machine service in order to preliminary evaluate a possibility of constructing a full-scale model and to train it appropriately.

N. Nikiforov (✉) · T. Tsoy · R. Safin · E. Magid
Laboratory of Intelligent Robotics Systems (LIRS), Institute of Information Technology and Intelligent Systems, Kazan Federal University, Kazan, Russia
URL: https://www.kpfu.ru/eng/itis/research/laboratory-of-intelligent-robotic-systems.com

T. Tsoy
e-mail: tt@it.kfu.ru

R. Safin
e-mail: safin.ramil@it.kfu.ru

E. Magid
e-mail: magid@it.kfu.ru

Y. Bai · M. Svinin
Information Science and Engineering Department, College of Information Science and Engineering, Ritsumeikan University, 1 -1 -1 Noji -higashi, Kusatsu, Shiga 525-8577, Japan
e-mail: yangbai@fc.ritsumei.ac.jp
URL:
http://www.en.ritsumei.ac.jp/academics/college-of-information-science-and-engineering/.com

M. Svinin
e-mail: svinin@fc.ritsumei.ac.jp

A. Ronzhin and V. Shishlakov (eds.), *Electromechanics and Robotics*, Smart Innovation, Systems and Technologies 232, https://doi.org/10.1007/978-981-16-2814-6_24

The obtained model demonstrated acceptable recognition rate. We also attempted to apply SURF and FLANN techniques for matching with the direct matching approach and the skeleton-based approach, but the matching results were not satisfactory.

1 Introduction

Nowadays, computer vision is used in various types of activities, including such complicated tasks as automatic object detection, search, and recognition. Recognition systems might concentrate on such objects as car plates [1], component labels [2], animals [3], and many others. In security field, human face recognition is used to ensure security of ATMs [4] and personal gadgets [5]. In the medical field, computer vision helps to detect leukocytes [6], cancer [7] and other diseases and disorders. In search and rescue operations, when it is physically dangerous for a person to operate within a dangerous zone, robots that replace people in victim search actively apply computer vision techniques that allow autonomous functions of a vehicle [8] and teleoperation remote control [9]. The list of objects of interest constantly replenishes as new problems arise that are difficult for human precise and fast detection and recognition capabilities.

Gesture recognition is often used as an important communication channel not only in human-human and human-animal interaction, but in human-robot interaction [10] and in a human communication with other intelligent devices and objects, for example, in operating household devices [11] or in interactive game-based learning [12]. In [13] authors demonstrated a hand gesture interface system for appliances' control in smart home environments [14]. In [15], a depth camera extracted a hand depth silhouette and the obtained images were recognized using a trained random forest[16]. In [17], a gesture recognition system generated an appropriate command, which allowed a selection, a mouse control [18], an exit and other additional functionalities.

In this paper, we focus on human-robot interaction using gestures [19]. Our long-term project's goal is to attempt enabling the car-like robot Avrora Unior (Fig. 1) control with user gestures, while these gestures are not completely predefined in advance [20]. To develop a gesture control concept for the Avrora Unior robot, we carefully studied and tested locomotion capabilities of the robot and created an exhaustive list of basic commands that are required to control the robot. A group of students were asked to provide their own unlimited gestures, which in their opinion would correspond to each command from the list. We attempted to use SURF [21] and FLANN [22] approaches to allow gesture recognition without constructing a skeleton of a user. This paper presents preliminary results of the pilot study.

The rest of the paper is organized as follows. Section 2 describes gesture control concept for the Avrora Unior robot. Section 3 explains process of collecting dataset and its characteristics. In Sect. 4 we present pilot studies. Finally, we conclude in Sect. 5.

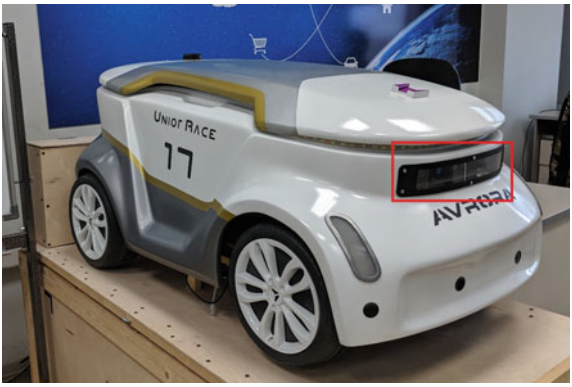

Fig. 1 The Avrora Unior robot. Red rectangular shows the Microsoft Kinect sensor

2 Gesture Control Concept

Use of gestures to interact with robots and smart devices has been explored in many works. Gestures allow controlling industrial robots via gesture-based user-friendly interface [23], Leap Motion technology [24] or Microsoft Kinect Controller [25]. Even a low-cost USB camera could successfully recognize and track user's hand movement and allow controlling simple activities [26, 27]. Phyo et.al. demonstrated an interaction with a humanoid NAO robot assistant with static hand gestures [28]. Gao et.al. presented a gesture-based smart wheelchair control for aged and disabled that was successfully validated within indoor environment [29]. Recently, Zhang et.al. demonstrated a gesture-based control of a real unmanned vehicle with Kinect-V2 sensor that employs upper body pose recognition for 13 joints and a dynamic time warping [30].

Our car-like robot Avrora Unior (Fig. 1) is equipped with the Microsoft Kinect sensor [31] that primarily targets for environment monitoring, mapping and obstacle avoidance. An important characteristic of this controller is the range at which it guarantees correct values. Considering the depth sensor, the maximum distance between an object and the sensor is limited to 3.5 m [32]. This limitation should be taken into account when control gestures are selected and when a dataset for machine learning of possible gestures is constructed [33].

The Avrora Unior robot could be controlled in teleoperational mode using a special one-hand held motion controller or locomote autonomously [34]. The primary purpose of introducing additional gesture control was to enable hands-free convenient testing process of new motion and interaction algorithms, which would allow to correct the robot movement and to avoid accidents in a much more robust and fast way. A user should approach the robot at a distance of 3–4 m and, while staying in front of the robot and entirely within the Microsoft Kinect sensor field of view, show a particular control gesture, which triggers a corresponding command execution. Currently, we selected control gestures to be static in order to reduce a complexity of their recognition.

Since for full-size autonomous vehicles safety is critical, typically all control gestures are predefined in advance and then the robot control system is taught to recognize these gestures using an exhaustive set of examples [35]. However, such approach requires an operator to carefully study the gestures and to be always concentrated in order to use a proper one. In our case would like to allow an unprepared user to control the robot intuitively, which implies an unconstrained control that could guess the user intentions and operate accordingly.

At the first step of the project, we created a broad list of remote control commands. The selection was based on several factors: a convenience of using a command, an importance of the command, and a possibility of using it remotely. A basic set contained just forward and backward motion, left and right turns. Next, several more complicated commands were added to the set, e.g., turning 180 degrees, approaching a static person and automatic parallel parking [36]. The set of the control commands is presented in Table 1.

3 Dataset

For dataset collection it was necessary to record a 30–40 s video with a static gesture for each command keeping a distance from a camera to a user within a range of 3–4 m, which matches the Kinect sensor capabilities [37]. A group of 17 people, students and employees of Laboratory of Intelligent Robotic Systems,[1] were asked to propose a gesture for each of the 18 control commands independently of the others. None of the participants knew what gestures the other participants had selected. This was done in order to analyze and select the most appropriate gesture for a particular command based on the received variety of gestures and their statistical distribution. Frames were extracted from each video using Matlab software. Table 1 presents a number of frames that were extracted from the collected set of 306 (17 people, 18 gestures each) short video sequences for each gestures.

Unfortunately, the similarity between gestures from different users was rather small. Only one command, "Full stop", had a small variety of patterns and a sigh similarity of gestures within a pattern in many cases (see Fig. 2).

4 Pilot Studies

4.1 Optimal Gestures Selection

A selection of optimal gestures for each command was based on 3 criteria: a size (of a bounding box for entire user's body, while demonstrating a command), an ease

[1] https://kpfu.ru/eng/itis/research/laboratory-of-intelligent-robotic-systems.

Table 1 The control commands list and a size of datasets (frames) for each command

No.	Command	Comments	Frames
1	Moving forward	At a predefined constant speed	30,755
2	Backward movement	At a predefined constant speed	29,011
3	Turn the wheels to the right	The wheels are turned, while the vehicle is static, the turning angle gradually increases	30,450
4	Turn the wheels to the left	The wheels are turned, while the vehicle is static, the turning angle gradually increases	28,990
5	Increase speed	Speed gradually increases by a predefined value	29,048
6	Decrease speed	Speed gradually decreases by a predefined value	28,829
7	Full stop	Emergency braking	30,233
8	A mode of ignoring a user is on	Ignore all commands, except for a full stop and a command to disable this mode	30,615
9	A mode of ignoring a user is off	Disable Ignore Mode	29,427
10	Move forward while turning the wheels to the right	With a predefined constant speed and a predefined angle of rotation of the wheels	28,781
11	Move forward while turning the wheels to the left	With a predefined constant speed and a predefined angle of rotation of the wheels	30,021
12	Move backward while turning the wheels to the right	With a predefined constant speed and a predefined angle of rotation of the wheels	26,581
13	Move backward while turning the wheels to the left	With a predefined constant speed and a predefined angle of rotation of the wheels	26,939
14	Automatic 180° turn		30,941
15	Automatic 90° turn to the right		29,691
16	Automatic 90° turn to the left		29,892
17	Drive to a user		31,205
18	Automatic parallel parking	The only complicated command within the set, which implies a closest parking spot search [38] and further parking [36]	31,847

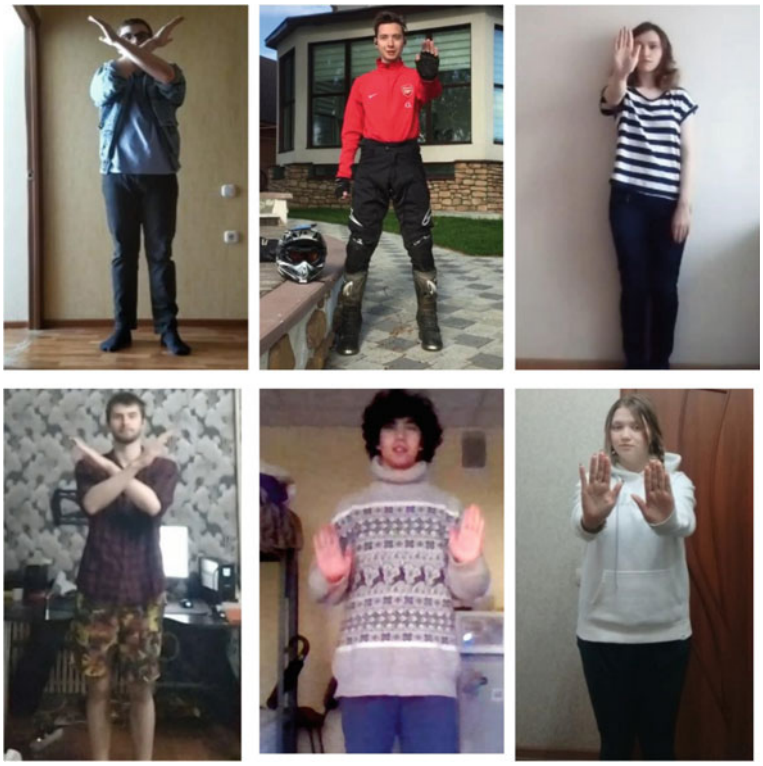

Fig. 2 Proposed gestures for "Full stop" command by two users

of use, and an absence of overlaps of upper limbs' links of a user's skeleton. It is important that a bounding box for a gesture does not exceed a predefined threshold, which depends on a distance from a user to the robot (in our pilot study selection we targeted for distance of 3–4 m) and the user's height; otherwise, the gesture might be out of field of view of the Kinect sensor, which is located at a height of 0.35 m from the ground [39]. For example, Fig. 3 presents unsuitable gestures that would be outside of field of view.

Some gestures might be difficult to perform due to physical limitations of an average human skeleton and joints' flexibility. This means it is highly likely that such gestures would be rarely used by a typical user. Figure 3 also presents an example of a gesture that might be difficult to repeat as this posture is rather uncomfortable for a typical human.

In order to analyze, a posture of a user and to extract a control signal from the posture a basic skeleton extraction approach was selected. To draw a basic skeleton, the OpenPose system [40] was used. It was decided to avoid using gestures that may not be recognized correctly due to overlapping upper limbs' links of a basic skeleton (see Fig. 4).

Fig. 3 Examples of unsuitable gestures due to their bounding box height (left and center) and uncomfortable posture (right)

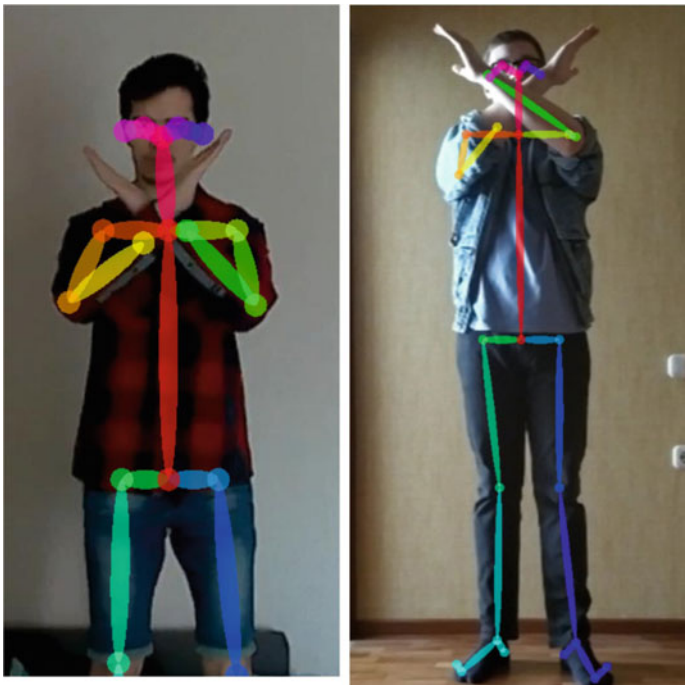

Fig. 4 Examples of unsuitable gestures due to overlapping of basic skeleton links

4.2 Training on Datasets

To verify if the constructed dataset of images would further allow us to create a full-scale model for machine learning, we used the Teachable machine service [41], which draws skeletons for dataset images and allows constructing and training a new model according to the selected training parameters. At this initial stage, for each command we used only 50 images from 5 participants to train a very basic model, which was taught to distinguish all 18 classes (all possible control commands of Table 1). Yet, even for a such small dataset the trained model demonstrated acceptable results (Fig. 5) that would obviously improve with the training set growth.

In order to check whether it is possible to detect correctly displayed gestures without constructing a skeleton and whether the skeleton use could improve the recognition precision we used Speeded-Up Robust Features (SURF [21]) and Fast Library for Approximate Nearest Neighbors (FLANN [22]) techniques. SURF was used to search for key features of the images (frames from the video sequences). Next, in order to obtain a quick and efficient matching, the comparison was performed using key feature comparator FLANN. The experiments obviously demonstrated that images with a similar posture of skeletons have more matches than in other cases. Figure 6 demonstrates an example of key feature points matching with two approaches: direct matching (the upper row in the figure; only 67 matches were successful) and using a basic skeleton (the central row in the figure; 103 matches were successful). Yet, for similar postures a skeleton-based approach also provided

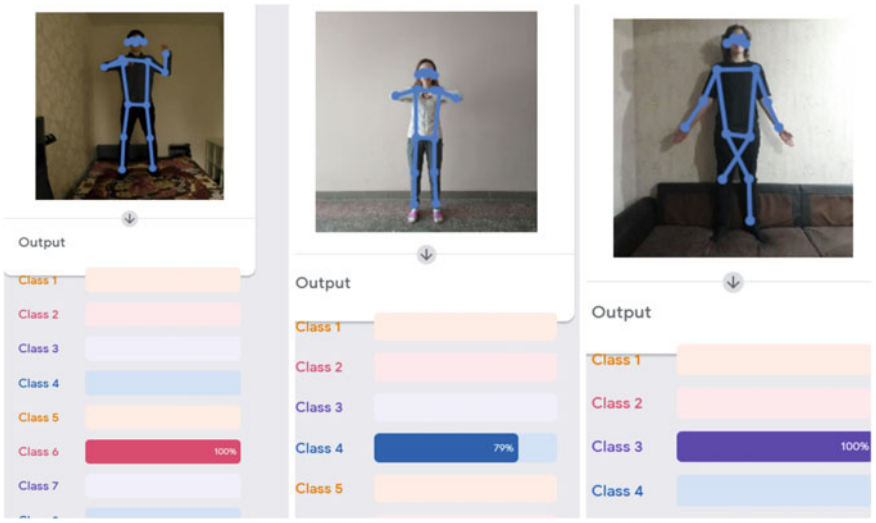

Fig. 5 An example of model training results using the Teachable machine service: 100% success for Class 6, 79% for Class 4, and 100% for Class 3. *Note* a mistake in the skeleton extraction for the image on the right

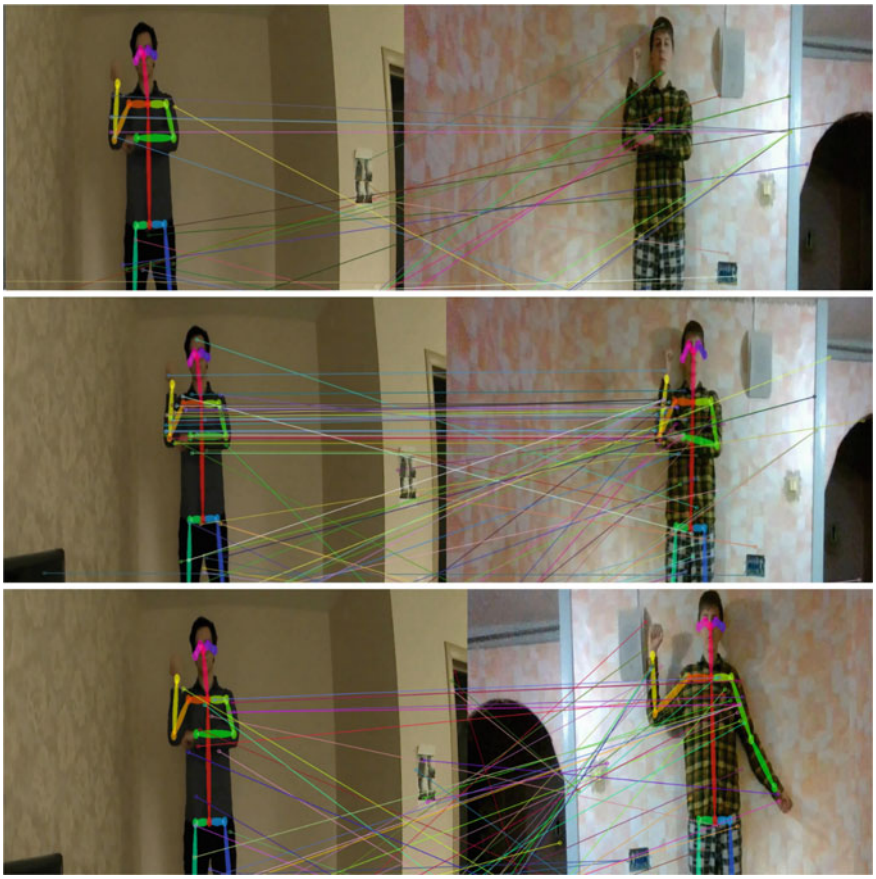

Fig. 6 A direct matching of postures without skeleton use (the upper row, 67 matches), a skeleton-based matching (the central row, 103 matches) and a skeleton-based matching of slightly different gesture (the bottom row, 91 matches)

a good level of matching (the bottom row in the figure; 91 matches were successful); however, in this example the two postures are objectively similar and thus a high level of matching features was appropriate.

Matching of some other images demonstrated that using a skeleton might not always cause a quantitative improvement relatively to the direct matching. For example, Fig. 7 demonstrates a particular case where the direct matching of postures without a skeleton quantitatively (technically) outperformed the skeleton-based matching (106 matches vs. 101 matches respectfully). Yet, a close look at the suggested corresponding matches demonstrates that in both cases a vast majority of the matches was wrong while the situation with proper matches was slightly better for the skeleton-based approach. The same issues with wrong matches applies to Fig. 6, but again the skeleton-based approach (Fig. 6, central row) demonstrated a significantly better

Fig. 7 A direct matching of postures without skeleton use (the upper row, 106 matches) and a skeleton-based matching (the bottom row, 101 matches)

amount of properly matched key features that the direct approach. The experimental results demonstrated that using SURF and FLANN techniques for matching, both with the direct approach and the skeleton-based approach, failed to provide acceptable level of matching for two similar human postures. Therefore, constructing a new classifier and teaching it appropriately becomes inevitable.

5 Conclusions

In this paper, we presented a concept for controlling the car-like robot Avrora Unior locomotion using human gestures. The list of 18 control commands contained basic and compound commands. A group of 17 volunteers used the commands' list to create individual control gestures independently. A small part of the obtained dataset of gestures (less than 0.2%) was used with the Teachable machine service in order to preliminary evaluate the possibility of constructing a full-scale model and to train it appropriately. The obtained model demonstrated acceptable recognition rate. We also attempted to apply SURF and FLANN techniques for matching with the direct matching approach and the skeleton-based approach, but they demonstrated an insufficient matching quality. Finally, we concluded that the collected dataset will allow constructing a good model that could be taught to successfully distinguish locomotion control gestures. The gestures from the developed datasets were not studied in any statistical aspects yet, which is a part of our ongoing work that will allow using the most statistically significant gestures and demonstrate reliability and validity of the proposed gesture selection criteria.

Acknowledgements This work was supported by the Russian Foundation for Basic Research (RFBR), project ID 19-58-70002. Forth and fifth authors acknowledge the support of the Japan Science and Technology Agency, the JST Strategic International Collaborative Research Program, Project No. 18065977.

References

1. Qadri, M.T., Asif, M.: Automatic number plate recognition system for vehicle identification using optical character recognition. In: 2009 International Conference on Education Technology and Computer, pp. 335–338. IEEE (2009)
2. He, L., Chao, Y., Suzuki, K., Wu, K.: Fast connected-component labeling. Pattern Recognit **42**(9), 1977–1987 (2009)
3. Nguyen, H., Maclagan, S.J., Nguyen, T.D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E.G., Phung, D.: Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In: 2017 IEEE international conference on data science and advanced Analytics (DSAA), pp. 40–49. IEEE (2017)
4. Ray, S., Das, S., Sen, A.: An intelligent vision system for monitoring security and surveillance of atm. In: 2015 Annual IEEE India Conference (INDICON), pp. 1–5. IEEE (2015)
5. Sutoyo, R., Harefa, J., Chowanda, A.: Unlock screen application design using face expression on android smartphone. In: MATEC Web of Conferences. vol. 54, p. 05001, EDP Sciences (2016)
6. Cuevas, E., Díaz, M., Manzanares, M., Zaldivar, D., Perez-Cisneros, M.: An improved computer vision method for white blood cells detection. Computational and Mathematical Methods in Medicine (2013)
7. Lee, H., Chen, Y.P.P.: Image based computer aided diagnosis system for cancer detection. Expert Syst Appl **42**(12), 5356–5365 (2015)
8. Al-Kaff, A., Moreno, F.M., de la Escalera, A., Armingol, J.M.: Intelligent vehicle for search, rescue and transportation purposes. In: 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), pp. 110–115. IEEE (2017)
9. Perez-Grau, F., Ragel, R., Caballero, F., Viguria, A., Ollero, A.: Semi-autonomous teleoperation of uavs in search and rescue scenarios. In: 2017 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1066–1074. IEEE (2017)
10. Shirwalkar, S., Singh, A., Sharma, K., Singh, N.: Telemanipulation of an industrial robotic arm using gesture recognition with kinect. In: 2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE), pp. 1–6. IEEE (2013)
11. Rashid, M., Han, X.: Gesture control of zigbee connected smart home internet of things. In: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 667–670. IEEE (2016)
12. Hsiao, H.S., Chen, J.C.: Using a gesture interactive game-based learning approach to improve preschool children's learning performance and motor skills. Comput Educat **95**, 151–162 (2016)
13. Rahman, A.M., Hossain, M.A., Parra, J., El Saddik, A.: Motion-path based gesture interaction with smart home services. In: Proceedings of the 17th ACM international conference on Multimedia, pp. 761–764 (2009)
14. Hussain, S., Schaffner, S., Moseychuck, D.: Applications of wireless sensor networks and rfid in a smart home environment. In: 2009 Seventh Annual Communication Networks and Services Research Conference, pp. 153–157. IEEE (2009)
15. Muñoz-Salinas, R., Medina-Carnicer, R., Madrid-Cuevas, F.J., Carmona-Poyato, A.: Depth silhouettes for gesture recognition. Pattern Recognit Lett **29**(3), 319–329 (2008)
16. Pal, M.: Random forest classifier for remote sensing classification. Int J Remote Sensing **26**(1), 217–222 (2005)

17. Rautaray, S.S.: Real time hand gesture recognition system for dynamic applications. Int J UbiComp (IJU) **3**(1) (2012)
18. Vivek Veeriah, J., Swaminathan, P.: Robust hand gesture recognition algorithm for simple mouse control. Int J Comput Commun Eng **2**(2), 219–221 (2013)
19. Galin, R., Meshcheryakov, R.: Review on human–robot interaction during collaboration in a shared workspace. In: International Conference on Interactive Collaborative Robotics, pp. 63–74. Springer (2019)
20. Malov, D., Edemskii, A., Saveliev, A.: Architecture of proactive localization service for cyber-physical system's users. In: International Conference on Interactive Collaborative Robotics, pp. 10–18. Springer (2019)
21. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput Vision Image Understanding **110**(3), 346–359 (2008)
22. Goel, A., Saxena, S.C., Bhanot, S.: Modified functional link artificial neural network. Int J Electri Comput Eng **1**(1), 22–30 (2006)
23. Tang, G., Webb, P.: The design and evaluation of an ergonomic contactless gesture control system for industrial robots. J Robotics (2018)
24. Chen, S., Ma, H., Yang, C., Fu, M.: Hand gesture based robot control system using leap motion. In: International Conference on Intelligent Robotics and Applications, pp. 581–591. Springer (2015)
25. Mikadlicki, K., Pajor, M.: Real-time gesture control of a CNC machine tool with the use Microsoft Kinect sensor. Int J Sci Eng Res **6**(9), 538–543 (2015)
26. Grif, H.S., Farcas, C.C.: Mouse cursor control system based on hand gesture. Procedia Technol **22**, 657–661 (2016)
27. Song, S., Yan, D., Xie, Y.: Design of control system based on hand gesture recognition. In: 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1–4. IEEE (2018)
28. Phyo, A.S., Fukuda, H., Lam, A., Kobayashi, Y., Kuno, Y.: A human-robot interaction system based on calling hand gestures. In: International Conference on Intelligent Computing, pp. 43–52. Springer (2019)
29. Gao, X., Shi, L., Wang, Q.: The design of robotic wheelchair control system based on hand gesture control for the disabled. In: 2017 International Conference on Robotics and Automation Sciences (ICRAS), pp. 30–34. IEEE (2017)
30. Zhang, B., Yang, M., Yuan, W., Wang, C., Wang, B.: A novel system for guiding unmanned vehicles based on human gesture recognition. In: 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR), pp. 345–350. IEEE (2020)
31. Zhang, Z.: Microsoft Kinect sensor and its effect. IEEE Multimedia **19**(2), 4–10 (2012)
32. Han, J., Shao, L., Xu, D., Shotton, J.: Enhanced computer vision with Microsoft Kinect sensor: a review. IEEE Transa Cybernet **43**(5), 1318–1334 (2013)
33. Safin, R., Lavrenov, R., Tsoy, T., Svinin, M., Magid, E.: Real-time video server implementation for a mobile robot. In: 2018 11th International Conference on Developments in eSystems Engineering (DeSE), pp. 180–185. IEEE (2018)
34. Magid, E., Lavrenov, R., Khasianov, A.: Modified spline-based path planning for autonomous ground vehicle. In: ICINCO (2), pp. 132–141 (2017)
35. Lavrenov, R., Zakiev, A.: Tool for 3d gazebo map construction from arbitrary images and laser scans. In: 2017 10th International Conference on Developments in eSystems Engineering (DeSE), pp. 256–261. IEEE (2017)
36. Imameev, D., Shabalina, K., Sagitov, A., Su, K.L., Magid, E.: Modelling Autonomous Parallel Parking Procedure for Car-Like Robot Avrora Unior in Gazebo Simulator, pp. 428–431 (2020)
37. Safin, R., Garipova, E., Lavrenov, R., Li, H., Svinin, M., Magid, E.: Hardware and software video encoding comparison. In: 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 924–929. IEEE (2020)
38. Imameev, D., Zakiev, A., Tsoy, T., Bai, Y., Svinin, M., Magid, E.: Lidar-based parking spot search algorithm. In: Thirteenth International Conference on Machine Vision. vol. 11605, p. 1160502. International Society for Optics and Photonics (2021)

39. Shabalina, K., Sagitov, A., Su, K.L., Hsia, K.H., Magid, E.: Avrora unior car-like robot in gazebo environment. In: International Conference on Artificial Life and Robotics, pp. 116–119 (2019)
40. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: Openpose: realtime multi-person 2d pose estimation using part affinity fields. IEEE Trans pattern Anal Mach Intell **43**(1), 172–186 (2019)
41. Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., Chen, A.: Teachable machine: Approachable web-based tool for exploring machine learning classification. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–8 (2020)