Society of Production Engineering

ICPES 2025

40th INTERNATIONAL CONFERENCE ON PRODUCTION ENGINEERING – SERBIA 2025

DOI: 10.46793/ICPES25.002M

University of Nis Faculty of Mechanical Engineering

Nis, Serbia, 18 - 19th September 2025

COGNITIVE MOBILE ROBOTICS BASED ON INTELLIGENT MECHANISMS OF LEARNING

Zoran MILJKOVIĆ[1*], Aleksandar JOKIĆ[1], Đorđe JEVTIĆ[1]

Orcid: 0000-0001-9706-6134; **Orcid:** 0000-0002-7417-4244; **Orcid:** 0000-0002-6917-1663

[1]University of Belgrade – Faculty of Mechanical Engineering, Serbia, Belgrade

*Corresponding author: zmiljkovic@mas.bg.ac.rs

Abstract: *The purpose of this contribution is the deployment of digital manufacturing through new cognitive intelligence mechanisms. With the implementation of Industry 4.0 principles, mobile intelligent robots utilized as transportation vehicles in the manufacturing system need a higher degree of autonomy to fulfill all the requirements of the contemporary market. Although industrial robots are common in manufacturing systems, mobile robotics requires the expertise of specialists in cognitive robotics issues to gain international competitiveness, particularly for small and medium-sized enterprises. The industrial mobile robots' autonomous subsystems based on deep machine learning provide significantly more flexibility as well as more accurate and robust real-time decisions compared to common deterministic sensor-based algorithms. The main goal of this paper is to create artificial intelligence-based solutions for cognitive mobile robotics within Industry 4.0 using a Machine Learning (ML) based approach, particularly deep learning (convolutional neural networks, recurrent neural networks, etc.). The focus of the paper is the generation of new ML-based cognitive intelligence mechanisms for obstacle avoidance, decision-making, and visual control of intelligent mobile robots, whereas the main goal of the paper is to demonstrate the possibility of integrating intelligent ML-based algorithms into a high-level cognitive architecture by enabling better understanding of the environment in real-time through the processing of higher-quality and more complex sensory data, thereby enhancing the overall flexibility of mobile robotic systems within intelligent manufacturing systems.*

Keywords: *mobile robots, intelligent control systems, digital manufacturing, cognitive intelligence mechanisms, deep learning, autonomous systems, visual servoing.*

1. INTRODUCTION

In the era of rapid development of technological innovation, the cognitive mobile robots have enormous potential to influence both the industrial and service sectors. In the industrial sector, intelligent mobile robots with learning mechanisms can support the declining workforce with repetitive, dull, and physically tiring tasks. However, for intelligent mobile robots to be effective in the manufacturing sector, many essential skills need to be acquired, such as highly accurate positioning [1], obstacle avoidance [2], perception [3], and decision-making [4]. One of the most promising research directions within the robotics domain that can fulfill the skills is the utilization of machine learning techniques. The limits and potential of deep machine learning within robotics systems are analyzed in great detail in [5]. The three primary axes of research development are learning, embodiment, and reasoning, with an aim to reach active learning, active manipulation, and integration of semantic and geometric reasoning. Active learning represents the process of learning through interaction with the environment or online learning, which differs from the standard learning pipeline where the model is trained and statically utilized in robotics application. Active manipulation is the process of adjusting robots pose, to increase the accuracy of the environment perception and the probability of the successful manipulation process. Integration of semantic and geometric reasoning includes the process of mapping and utilizing not only the geometric properties of the environment but also the human-centered meaning of the objects and their semantic relationships. Two primary approaches to achieving the aforementioned skills include the development of individual skills and their integration trough cognitive high-level architectures such as, subsampling architecture [6], cognitive maps [7], and Soar [8], or utilizing novel deep learning based Vision-Language-Action (VLA) models [9]. The main primary deficiency of the VLA models is in their lack of reliability stemming form their black box nature and their sensitivity to environmental factors. Therefore, in this paper we will analyze the three developed deep machine learning-based algorithm utilized to achieve the necessary skills for the implementation of cognitive mobile robots within the manufacturing environment.

2. SEMANTIC SEGMENTATION BASED OBSTACLE AVOIDANCE

For a mobile robot to work autonomously in a dynamic and unstructured industrial environment, it needs to be able to avoid different obstacles in its path. Therefore, a globally optimally planned path needs to have local minor deviations necessary to avoid obstacles detected with the robot's sensors.

In this section, we analyze one of the approaches for obstacle avoidance based on semantic segmentation [2]. The custom version of the ResNet model is trained on the SunRGBD indoor dataset [10]. Since the experimental evaluation is performed on the mobile robot RAICO (Robot with Artificial Intelligence based COgnition), which has the Nvidia Jetson Nano development board, the utilized convolutional neural network needed to be lightweight. The state machine utilized for the obstacle avoidance algorithm is shown in Fig. 1.

The path the mobile robot takes during the obstacle avoidance process can be seen in Fig. 2. The obstacle detection is performed by analyzing semantic maps and checking if there is something other than class "floor" in the mobile robot's path. An example of one obstacle avoidance procedure, shown from RAICO's perspective, is illustrated in Fig. 3. As can be seen, the floor is labeled with red, yellow represents the obstacle, black denotes the tables, and teal represents the walls. Images generated by the mobile robot and the transformation of the image space occupancy grid defined with a semantic map to a 2D occupancy grid in the mobile robot frame are shown in Fig. 4. Green represents the free space, while red represents the occupied grids.

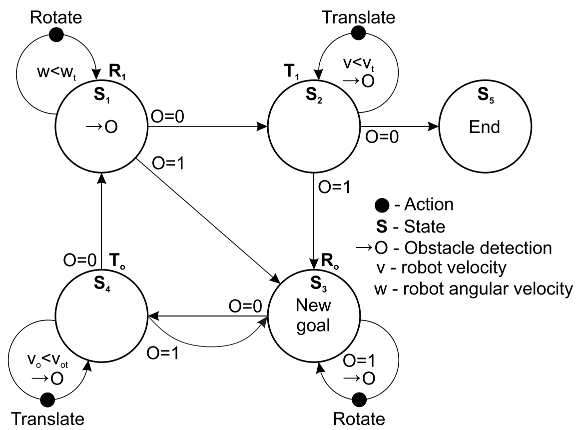

Figure 1. State machine for obstacle avoidance algorithm.

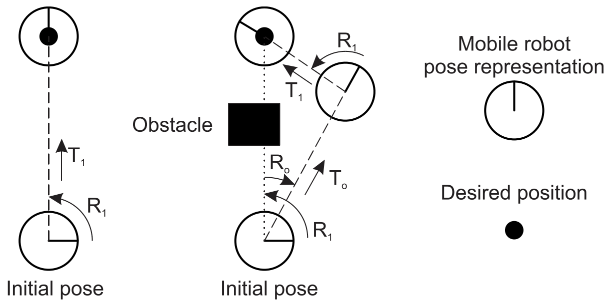

Figure 2. Mobile robot during path following (left) and during obstacle avoidance algorithm (right).

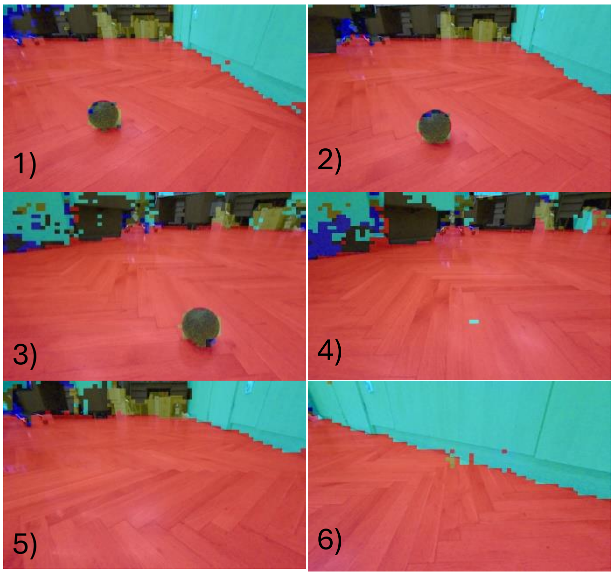

Figure 3. Semantic segmentation of the mobile robot's scene.

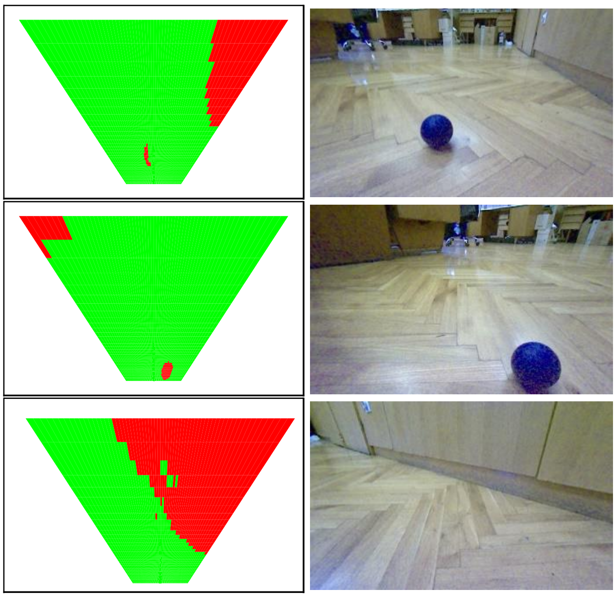

Figure 4. Visualization of the occupancy grid generated from the visual system.

The path the mobile robot took during the experimental evaluation can be seen in Fig. 5.

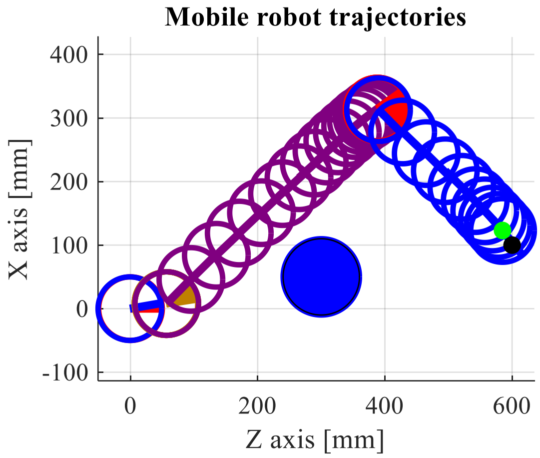

Figure 5. Mobile robot path.

3. SEMANTIC SEGMENTATION-BASED VISUAL SERVOING

One of the key requirements for the efficient utilization of cognitive architectures is to reuse the data for multiple behaviors. Having that in mind, the second behavior considers accurate positioning of the mobile robot, based on semantic stereo visual servoing. Since the standard odometry-based mobile robot positioning system cannot provide high levels of accuracy, needed for the manipulation tasks in the manufacturing environment, the stereo visual servoing based on semantic segmentation is employed [1].

The images generated in the laboratory manufacturing environment are converted into semantic maps by utilizing a convolutional neural network (Fig. 6). The dataset comprises five classes: four machine tool classes and one background class.

Afterwards, semantic maps generated in the desired and current pose (position and orientation) are registered [11]. The registration represents the process of image transformation with an objective function to align two images. The semantic maps registration can be seen in Fig. 7 for both images of the stereo vision system. Combined semantic maps of two images (current and desired) are shown in the first row. One of the initially generated semantic maps is shown with shades of pink, while the other is shown with green. After the registration process is completed, the images are displayed in the second row. If the green and pink elements are aligned accurately, they are displayed in gray.

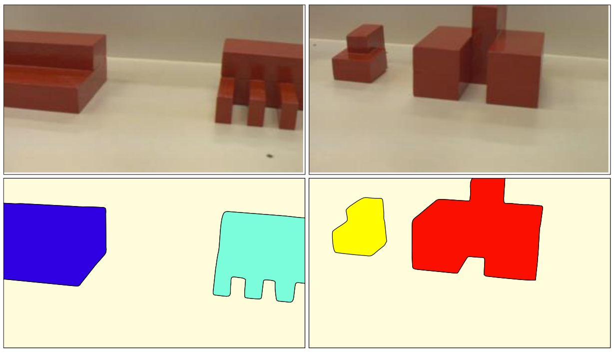

Figure 6. Images and semantic maps generated from the stereo visual system.

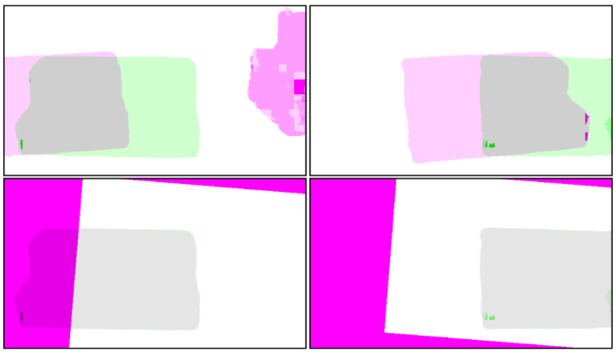

Figure 7. Semantic map registration process.

The output of the image registration process is the transformation matrix $\mathbf{T} \in \mathbb{R}^{3\times3}$. That transformation matrix is utilized to align all the control pixels belonging to the same machine in the current $\mathbf{c}_c = (u_c, v_c, 1)^T$ and desired $\mathbf{c}_d = (u_d, v_d, 1)^T$ image (1):

$$\mathbf{c}_c = \mathbf{T}\mathbf{c}_d. \tag{1}$$

After the registration is performed and the pixels are connected by utilizing the transformation matrix, the difference between those pixels can be used to determine the necessary velocities in the camera frame (2):

$$\dot{\mathbf{c}} = \mathbf{J}\mathbf{v}_{cam}, \tag{2}$$

where $\dot{\mathbf{c}}$ is the difference between current and desired control pixels, $\mathbf{J}$ is the image Jacobian matrix [12], and $\mathbf{V}_{cam}$ are the velocities generated in the camera frame. Finally, the velocities in the mobile robot frame ($\mathbf{V}_{mr} \in \mathbb{R}^{6\times1}$) are generated by utilizing a visual servoing controller for k defined control pixels (3):

$$\mathbf{v}_{mr} = \lambda \begin{pmatrix} \mathbf{J}_1\mathbf{M} \\ \vdots \\ \mathbf{J}_k\mathbf{M} \end{pmatrix}^{-1} \begin{pmatrix} \dot{\mathbf{c}}_1 \\ \vdots \\ \dot{\mathbf{c}}_k \end{pmatrix}, \tag{3}$$

where λ is the gain value, and $\mathbf{M} \in \mathbb{R}^{6\times6}$ is the velocity transformation matrix from the camera to the mobile robot frame.

Both simulation and real-world experimental evaluations with the semantic segmentation-based stereo visual servoing algorithm show high levels of accuracy [1]. Initial pose of mobile robot RAICO in simulated and real-world experiment can be seen in Fig. 8. The path mobile robot took during the first experiment can be seen in Fig. 9. The accuracy of the final pose can also be shown by visualizing the absolute pixel difference between the final and the desired images (Fig. 10).

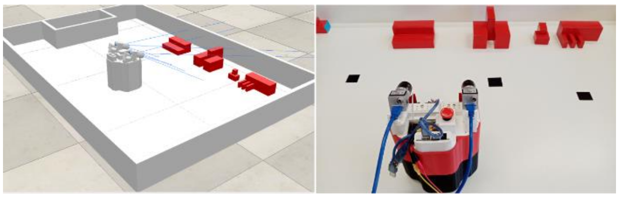

Figure 8. Initial pose of mobile robot RAICO in the simulation (left) and real-world (right) experimental setup.

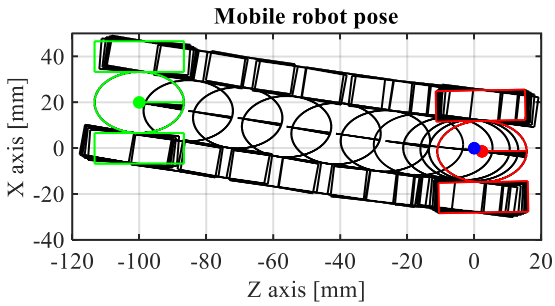

Figure 9. The path RAICO took from starting (green), intermediate (black), to the final (red) pose. The desired position is shown with a blue dot.

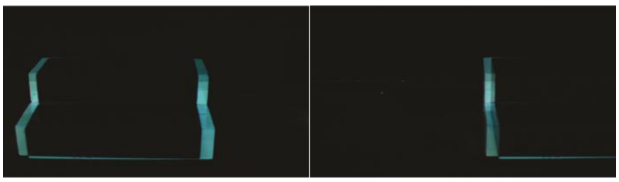

Figure 10. Absolute pixel difference between the desired and final images.

The black color represents a low difference between pixels, while light blue represents the parts of the image with noticeable differences. An additional important feature of the semantic segmentation-based stereo visual servoing is the ability to utilize the images generated in simulation for visual servoing, which is provided in the last experimental evaluation [1]. The desired images (Fig. 11) are generated in simulation and utilized in the real-world environment.

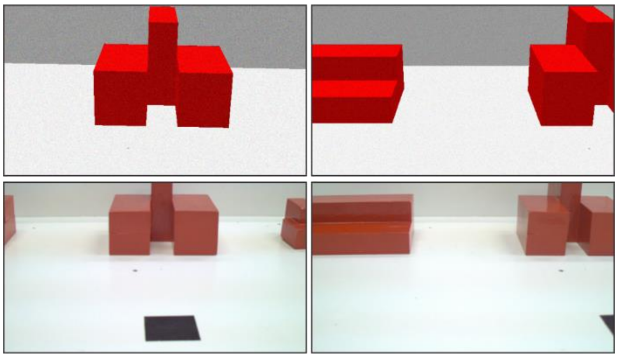

Figure 11. Desired images generated in simulation (first row) and real-world current images (second row).

The mobile robot managed to achieve the desired pose, which opens a possibility to change the desired image adaptably, significantly increasing the flexibility of the visual servoing algorithm.

4. ACCURACY ESTIMATION SYSTEM BASED ON DEEP LEARNING

The last system that utilizes deep learning in mobile robot applications is the accuracy estimation system that predicts the probability of a successful pickup procedure [4]. The input variables include the pose of the mobile robot, as well as the image of the final scene before the pickup procedure is initiated. The standard ResNet18 model (Fig. 12) is adapted and trained on the custom dataset to predict if the pickup procedure will be successful with a standard softmax output function.

The output of the model is utilized to assess if the current pose of the mobile robot is adequate for the pickup procedure.

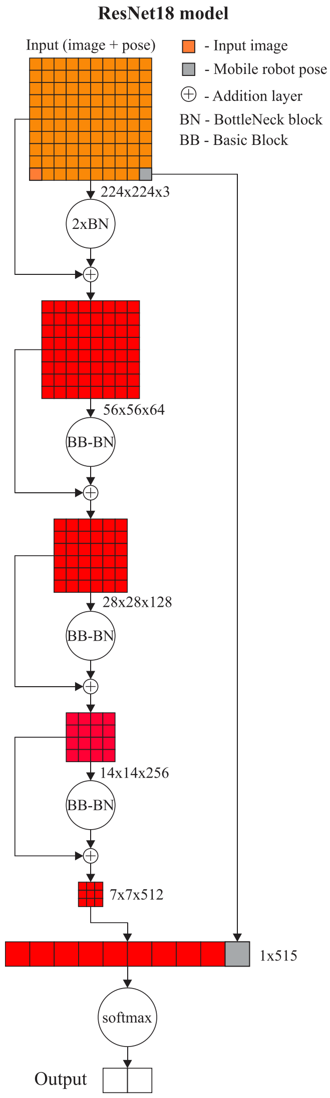

Figure 12. ResNet18 model for the probability of the pickup success prediction.

The experimental evaluation showed that high confidence of the model is highly correlated with the gripping success.

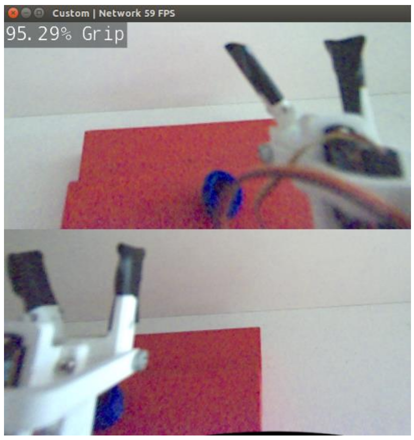

Figure 13. Mobile robot's view of the scene and gripping accuracy prediction.

However, the overall accuracy of the model is 70%. Figures 13 and 14 show the mobile robot performing the pickup procedure from the mobile robot's perspective and from the side view, respectively.

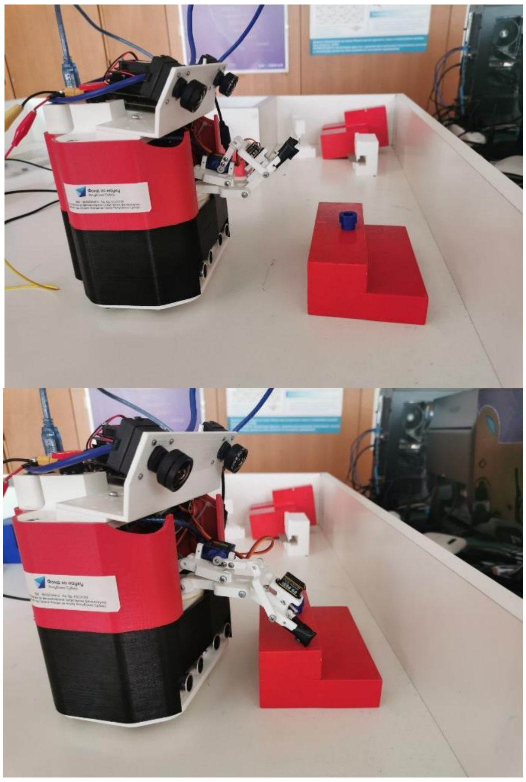

Figure 14. Side view of mobile robot RAICO during the gripping procedure.

5. CONCLUSION

In this paper, we have presented the results regarding the utilization of the intelligent learning mechanisms within the cognitive mobile robot domain. Three primary subsystems utilized to demonstrate the advantages of the deep machine learning approaches are obstacle avoidance, visual servoing, and prediction of the accuracy of pickup operation. The obstacle avoidance is performed by utilizing semantic segmentation of the scene, where the mobile robot RAICO can determine the difference between the classes of objects in its environment and the floor class. After successful semantic segmentation, the image is projected to a 2D occupancy grid and utilized within the obstacle avoidance procedure. The second algorithm regards the accurate positioning of the mobile robot. For that, we utilized stereo visual servoing based on semantic segmentation. After the registration of semantic maps, it is possible to determine the velocities the mobile robot needs to achieve to position itself accurately. Lastly, the standard ResNet18 model is utilized in the image classification problem to determine if the current pose is adequate for the pickup procedure. Mobile robot pose and image information represent the input into the desired model. The future research direction includes the integration of the developed algorithm into a cognitive control architecture.

ACKNOWLEDGEMENT

This research was supported by the Science Fund of the Republic of Serbia, grant No. 17801, Cybersecurity of Motion Control Systems in Industry 4.0 – MCSecurity, and by the Ministry of Science, Technological Development and Innovations of the Serbian Government under the contract No. 451-03-137/2025-03/200105.

REFERENCES

[1] A. Jokić, M. Petrović, Z. Miljković, Semantic segmentation based stereo visual servoing of nonholonomic mobile robot in intelligent manufacturing environment, Expert Systems with Applications, Vol. 190, p. 116203, 2022.

[2] M. Petrović, Z. Miljković, A. Jokić, Efficient Machine Learning of Mobile Robotic Systems based on Convolutional Neural Networks, in Artificial intelligence for Robotics and Autonomous Systems. Studies in Computational Intelligence, A. Koubaa and A. T. Azar, Eds., Springer, 2023, pp. 1–26.

[3] A. Jokić, M. Petrović, Z. Miljković, Real-time Mobile Robot Perception based on Deep Learning Detection Model, in New Technologies, Development and Application V (NT 2022). Lecture Notes in Networks and Systems, Springer, Cham, 2022, pp. 670–677.

[4] A. Jokić, M. Petrović, Z. Miljković, Mobile robot decision-making system based on deep machine learning, in 9th International

Conference on Electrical, Electronics and Computing Engineering (IcETRAN 2022), 2022, pp. 653–656.

[5] N. Sünderhauf et al., The limits and potentials of deep learning for robotics, International Journal of Robotics Research, Vol. 37, No. 4–5, pp. 405–420, 2018.

[6] R. Brooks, A robust layered control system for a mobile robot, IEEE journal on robotics and automation, Vol. 2, No. 1, pp. 14–23, 2003.

[7] S. Vasudevan, S. Gächter, V. Nguyen, R. Siegwart, Cognitive maps for mobile robots—an object based approach, Robotics and Autonomous Systems, Vol. 55, No. 5, pp. 359–371, 2007.

[8] J. E. Laird, The Soar cognitive architecture. MIT press, 2019.

[9] G. R. Team et al., Gemini robotics: Bringing ai into the physical world, arXiv preprint arXiv:2503.20020, 2025.

[10] S. Song, S. P. Lichtenberg, J. Xiao, SUN RGB-D: A RGB-D scene understanding benchmark suite, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 567–576.

[11] Z. Miljković, A. Jokić, M. Petrović, Image Registration Algorithm for Deep Learning-Based Stereo Visual Control of Mobile Robots, in Deep Learning for Unmanned Systems. Studies in Computational Intelligence, A. Koubaa and A. T. Azar, Eds., Springer, Cham, 2021, pp. 447–479.

[12] F. Chaumette, S. Hutchinson, Visual servo control. I. Basic approaches, IEEE Robotics and Automation Magazine, Vol. 13, No. 4, pp. 82–90, 2006.