DOI 10.52209/1609-1825_2026_1_32 UDC 004.896:681.518.3

A ROS 2 Based Modular Control Architecture Integrating Visual Servoing and Model Predictive Control for Robotic Marking

*[1,2]**TANATOV Aibar,** Doctoral Student, jazbalar@gmail.com,*
*[1,3]**OMAROV Batyrkhan,** PhD, Professor, batyahan@gmail.com,*
*[1,4]***KUATOVA Moldir,** PhD, Professor, kuatova.moldyr@gmail.com,*
*[5]**MENDES Daniel Almeida Mateus,** PhD, Adjunct Professor, mmendes@isr.uc.pt,*
*[1]**ABEKOVA Aidana,** Doctoral Student, Abekova310792@gmail.com,*
[1]Institute of Mechanics and Engineering named after academician U.A. Dzholdasbekov, 29 Shevchenko Street, Almaty, Kazakhstan,
[2]NPJSC «Al-Farabi Kazakh National University», 71 Al-Farabi Avenue, Almaty, Kazakhstan,
[3]JSC «International Information Technology University», 34/1 Manas Street, Almaty, Kazakhstan,
[4]LLP «International Engineering and Technological University», 89/21 Al-Farabi Avenue, Almaty, Kazakhstan,
[5]Polytechnic University of Coimbra, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, Coimbra, Portugal,
**corresponding author.*

Abstract. *This paper presents a ROS 2-based intelligent control architecture for high-precision autonomous marking. The system integrates model predictive control (MPC) with dual-camera visual servoing to achieve submillimeter accuracy. A layered control stack combines global perception from an overhead camera with local correction from a tool-mounted camera, enabling accurate six-DoF pose estimation through fiducial markers and hand-eye calibration. Trajectory planning is formulated as a constrained optimization in CasADi, supporting real-time motion generation under workspace and actuator limits. Implemented on a 6-DOF collaborative arm with STM32-based RS-485 motor controllers, the system achieves final errors of 0.7 mm and closed-loop frequencies near 8 Hz. Results demonstrate the viability of combining vision and optimal control in a modular ROS 2 framework for industrial automation.*

Keywords: *ROS 2, visual servoing, model predictive control, robotic marking, dual-camera feedback, STM32 microcontroller, trajectory optimization, hand-eye calibration, real-time control, precision robotics.*

Introduction

Modern industrial environments require robots to perform precision tasks such as marking or engraving under variable, partially unstructured conditions – for example, engraving IDs on nameplates placed in random positions and orientations on a workbench. Traditional fixed automation lines typically rely on pre-defined trajectories and purely kinematic, position-controlled motion. Such architectures assume repeatable fixturing and accurate offline calibration; they work well for rigid, highly structured production settings, but they struggle when workpieces are displaced, rotated, or replaced by different geometries. Maintaining accuracy then often requires manual intervention, re-teaching trajectories, or conservative safety margins, which reduces throughput and flexibility [1-2].

Conventional industrial marking systems usually combine a single camera (if any) with simple position or velocity controllers running on proprietary hardware. Vision, when present, is often used only for coarse localization or post-process inspection rather than being deeply integrated into the control loop. As a result, these systems are sensitive to calibration errors and cannot easily adapt to changing plate positions, surface conditions, or lighting. Moreover, monolithic control architectures complicate upgrades: adding a new sensor or trying a different control algorithm often requires non-trivial re-engineering of the entire software stack [3].

Model Predictive Control (MPC) offers a principled way to address such variability. MPC plans optimal actions over a finite horizon while enforcing system constraints and predicting future behavior to adjust commands accordingly [1]. Prior studies have shown that MPC can handle multivariate constraints and improve manipulator trajectory smoothness and tracking under changing conditions, especially when dynamics or limits play a crucial role [2]. Visual servoing complements MPC by providing spatial feedback from cameras to adjust the robot pose in real time. Position-based visual servoing, which depends on accurate pose estimation, is robust for precision tasks such as alignment or marking [3]. A dual-camera setup – an overhead camera for global guidance and a tool-mounted (flange) camera for local refinement – further enhances robustness: the overhead view brings the robot near the target, while the flange camera supplies high-resolution corrections of small residual errors, enabling sub-millimeter precision [4].

Reinforcement learning (RL) can further improve adaptability by learning control policies from interaction. Hybrid MPC+RL approaches are promising for tasks that demand both precision and flexibility, especially when fiducial markers are absent or the environment changes significantly [8-9].

On the hardware side, embedded platforms such as STM32 microcontrollers enable low-latency motor control, while communication protocols like CAN and RS-485 provide real-time data exchange with actuators. The ROS 2 middleware adds a modular, distributed framework to orchestrate perception, planning, and control, using DDS for deterministic message passing [5-6]. Precise calibration methods – such as ArUco-based pose estimation, hand-eye calibration, and consistent transform trees across robot, camera, and workpiece frames – are critical to achieve sub-millimeter accuracy when closing the loop between vision and motion [7].

The aim of this work is to develop and experimentally validate a modular ROS 2-based control architecture for high-precision robotic marking that tightly integrates dual-camera visual servoing with model predictive control. The proposed architecture (i) combines global guidance from an overhead camera with local refinement from a flange-mounted camera, (ii) uses an MPC planner formulated in CasADi to generate smooth, constraint-aware trajectories in real time, and (iii) is implemented on a 6-DoF collaborative arm with STM32-based RS-485 motor controllers. Experiments on randomly placed plates demonstrate sub-millimeter positioning accuracy and a roughly seven-fold reduction in final positioning error compared to a single-camera baseline, while maintaining real-time closed-loop operation. Section 2 describes the system design and implementation, Section 3 reports experimental validation and quantitative comparison with a baseline controller, and Section 4 concludes with insights and directions for future work.

Methodology

The proposed control architecture is realized as a set of ROS 2 nodes for perception, planning, and control. Figure 1 outlines the control stack and the data flow from the vision sensors to the robot actuators [10]. An overhead camera observes the workpiece and estimates a coarse target pose in the workspace frame, guiding the robot toward the plate. As the robot approaches, a flange-mounted camera detects fiducial markers on the plate and computes a precise six-DoF pose, providing local refinement. This two-stage visual feedback – global then local – enables accurate target localization even in the presence of large initial errors [11]. An MPC-based planner transforms these pose estimates into optimized trajectories under joint and workspace constraints, while low-level motor controllers execute the resulting setpoints. ROS 2 ensures deterministic, real-time communication between all modules, making the architecture modular and extensible.

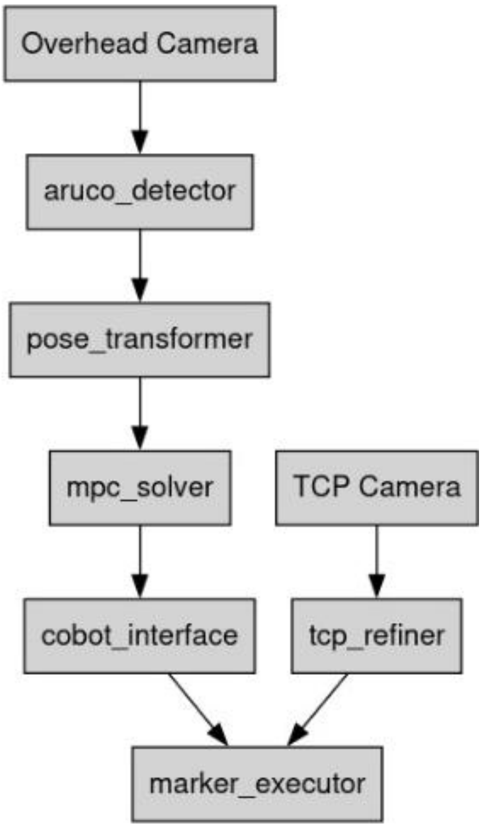


Figure 1 – ROS 2 node graph showing perception, planning, and control modules with data flow

A key element is the transform tree linking all coordinate frames: robot joints, cameras, workpiece, and tool. Static and dynamic transforms align the world, base, end-effector, cameras, and target marker. After hand-eye calibration, the flange camera-tool transform is fixed, even during motion. Using this calibrated tree, the overhead camera’s coarse plate pose is mapped to the base frame for approach, while the flange camera’s refined pose is mapped to the tool frame for final alignment [12][13]. The world-to-marker transform is estimated online from vision, updating the target pose with each camera feedback cycle.

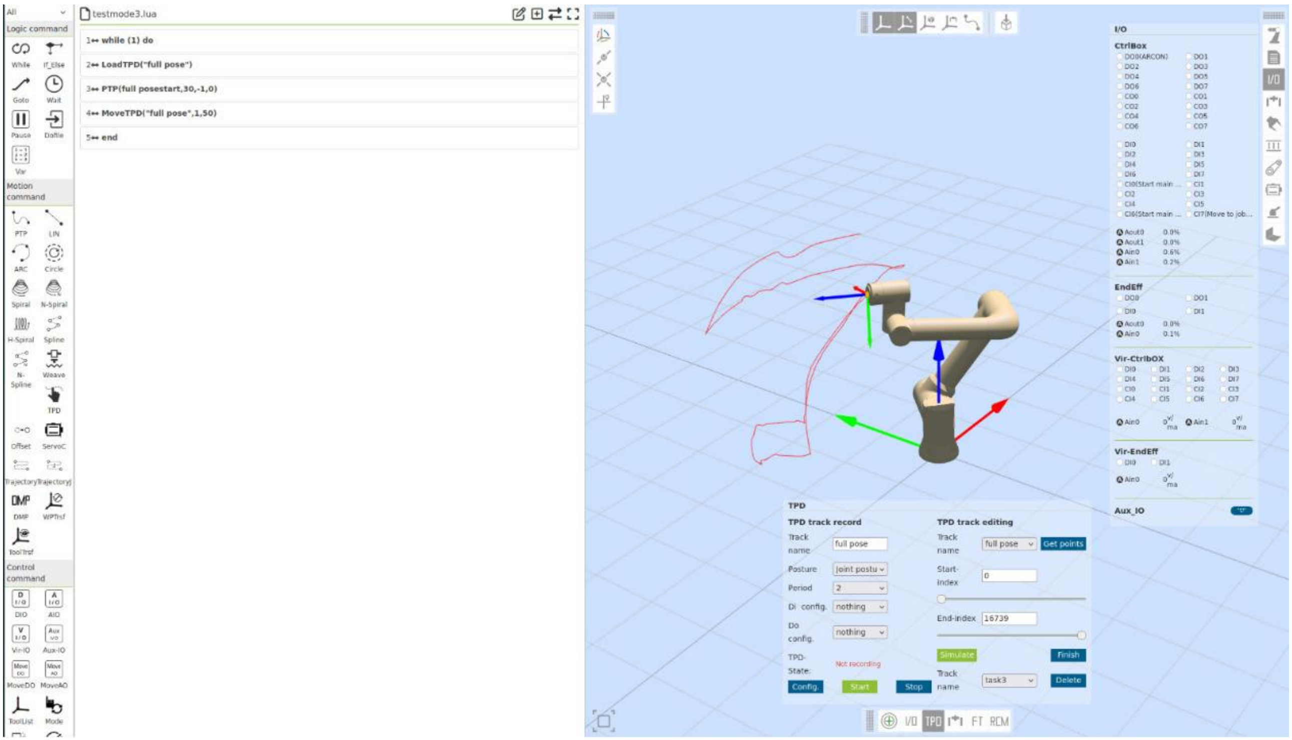
Robot motion is governed by MPC to enforce both accuracy and smoothness. The planner, implemented in CasADi, formulates and solves a constrained optimization at a fixed control period. It minimizes a cost that penalizes tracking error and control effort subject to joint limits, workspace bounds, and collision-avoidance constraints [14][15]. The solution is a short-horizon sequence of joint setpoints, which a low-level controller tracks. Early point-to-point (PTP) motions at full speed caused high jerk and vibration, so we adopted time-parameterized trajectories: velocities ramp smoothly within MPC constraints and are updated every control cycle, typically with a reduced peak speed to respect limits while maintaining throughput. Compared to a 100% PTP script, this markedly reduces discontinuities, jerk, and acceleration spikes, improving engraving quality and reducing mechanical stress.

Results

We evaluated the system in robotic marking trials, focusing on positioning accuracy, trajectory smoothness, computational performance, convergence, and robustness. In each trial, the robot engraved text on randomly placed plates; each run used five plates with different placements, and multiple runs ensured statistical significance. Distances are in millimeters (mm) and times in seconds (s).

We evaluated the system in robotic marking trials, focusing on positioning accuracy, trajectory smoothness, computational performance, convergence, and robustness. In each trial, the robot engraved text on a randomly placed plate; each run used five plates with different placements, and multiple runs ensured statistical significance. Distances are in millimeters (mm) and times in seconds (s).

To quantify the benefit of the proposed dual-camera + MPC architecture, we compared it against a baseline configuration that uses only the overhead camera for visual guidance while keeping the same ROS 2 infrastructure and low-level controllers. In the baseline, the flange-mounted camera and fine-alignment stage were disabled, so the robot executed the engraving directly after the coarse approach.


testmode3.lua

```
while (1) do
LoadTPD("full pose")
PTP(full posestart,30,-1,0)
MoveTPD("full pose",1,50)
end
```

TPD

TPD track record

Track name: full pose

Posture: joint postu

Period: 2

DI config.: nothing

Do config.: nothing

TPD-State: Not recording

TPD track editing

Track name: full pose

Start-index: 0

End-index: 16739

Track name: task3

Figure 2 – Summarizes the final positioning error over all runs: the dual-camera + MPC architecture reduces the error from about 5.8 mm to 0.8 mm

This single-camera baseline produced a final average positioning error at the tool-tip of about 5.8 mm, and in several trials the engraved text was visibly offset or partially outside the intended region. In contrast, the dual-camera configuration with MPC consistently achieved sub-millimeter accuracy, with a median final error of ~0.8 mm and a standard deviation of ~0.2 mm across plates, corresponding to a roughly seven-fold reduction in error at comparable motion speeds. No systematic drift was observed; the first and last plates in long runs exhibited statistically similar errors.

The temporal behavior of the fine-alignment phase is characterized by a rapid drop in error from several millimeters after the coarse approach to below 1 mm within ~0.4 s, settling around 0.7 mm after ~0.6 s. The MPC loop runs at 7-9 Hz, with typical solve times well below the control period, and no control deadlines were missed. Thus, the dual-camera + MPC architecture improves accuracy by an order of magnitude while maintaining real-time performance and adding only minor overhead to the engraving cycle.

Using the dual-camera + MPC approach, the system consistently achieved sub-millimeter precision, with a median final error of ~0.8 mm (best cases 0.7 mm). Despite low-cost hardware and unstructured plate placement, all engravings stayed within the intended boundaries; discrepancies were only instrument-detectable. Initial placement errors of 5-10 mm from a preset approach pose were first reduced to ~2-3 mm by the overhead camera, then brought below 1 mm by the flange-mounted camera with MPC refinement. This sequential coarse-to-fine strategy was essential: the overhead camera alone could not reliably surpass ~2 mm, while a flange camera alone would struggle with large initial offsets. Performance was highly repeatable, with a standard deviation of ~0.2 mm and no observable drift across sequential engravings. Overall, the dual-camera configuration achieved ~0.8 mm versus ~5.8 mm for the single-camera baseline, underscoring the importance of local visual feedback with MPC for reliable sub-millimeter marking in unstructured conditions.

Beyond accuracy, trajectory smoothness is vital for engraving quality and mechanical longevity. Compared to rapid point-to-point (100% speed) motion with sharp velocity changes and high jerk, MPC-planned trajectories (≈50% speed) ramp velocity gradually, reducing peak acceleration, avoiding overshoot or oscillation, and keeping the tool stable on contact. This smoother motion lowers mechanical stress on the manipulator while still reaching the target promptly. Most of each control cycle remained idle and even in the worst case (~15 ms solve time) the loop only briefly dropped to ~7 Hz, confirming real-time operation with comfortable computational headroom.

We also analyzed convergence to the final pose during fine alignment. From the coarse approach position, the initial error is several millimeters; in a representative trial, the error fell below 1 mm within ~0.4 s and settled around 0.7 mm after ~0.6 s. The curve exhibited two phases: a rapid initial drop as flange-camera feedback removed most misalignment in the first half second, followed by a slower refinement correcting the last fraction of a millimeter. Across runs, the convergence profile remained consistent, with similar time constants and variance. In practice, the robot achieved sub-millimeter alignment very quickly once the plate was visible to the flange camera, so fine correction added little overhead and kept the overall cycle time from detection to engraving low.

System robustness was evaluated through both baseline and extended-run tests. In the baseline configuration (flange camera disabled), the robot's final average error increased to ~5-6 mm, and in some trials the engraved text was visibly offset or partially missing the intended area. The single-camera approach therefore could not reliably guarantee proper alignment. In contrast, the full dual-camera system kept the error under 1 mm and all engravings were correctly placed, highlighting the necessity of two-tier visual feedback for high precision.

For long-term robustness, we engraved 15 plates in a row (cycling through five randomly placed plates) without any recalibration or manual intervention; accuracy did not degrade, and no cumulative drift was observed. The system also remained stable under slight lighting variations, with ArUco detection unaffected, indicating that the architecture can maintain calibration and accuracy over extended operation in realistic industrial conditions.

Conclusions

This study demonstrated a ROS 2-based modular control architecture for high-precision robotic marking that combines MPC, dual-camera visual servoing, and embedded actuation. The overhead camera provides global guidance, while the flange-mounted camera supplies local refinement. An MPC planner fuses these inputs to generate optimized motion, enabling consistent sub-millimeter positioning under dynamic and uncertain conditions. Closed-loop frequencies of 7-9 Hz with low-latency feedback satisfied real-time requirements, and the integration of STM32 microcontrollers, high-speed serial communication, and ROS 2 proved effective for deterministic coordination. The architecture is extensible and modular, allowing

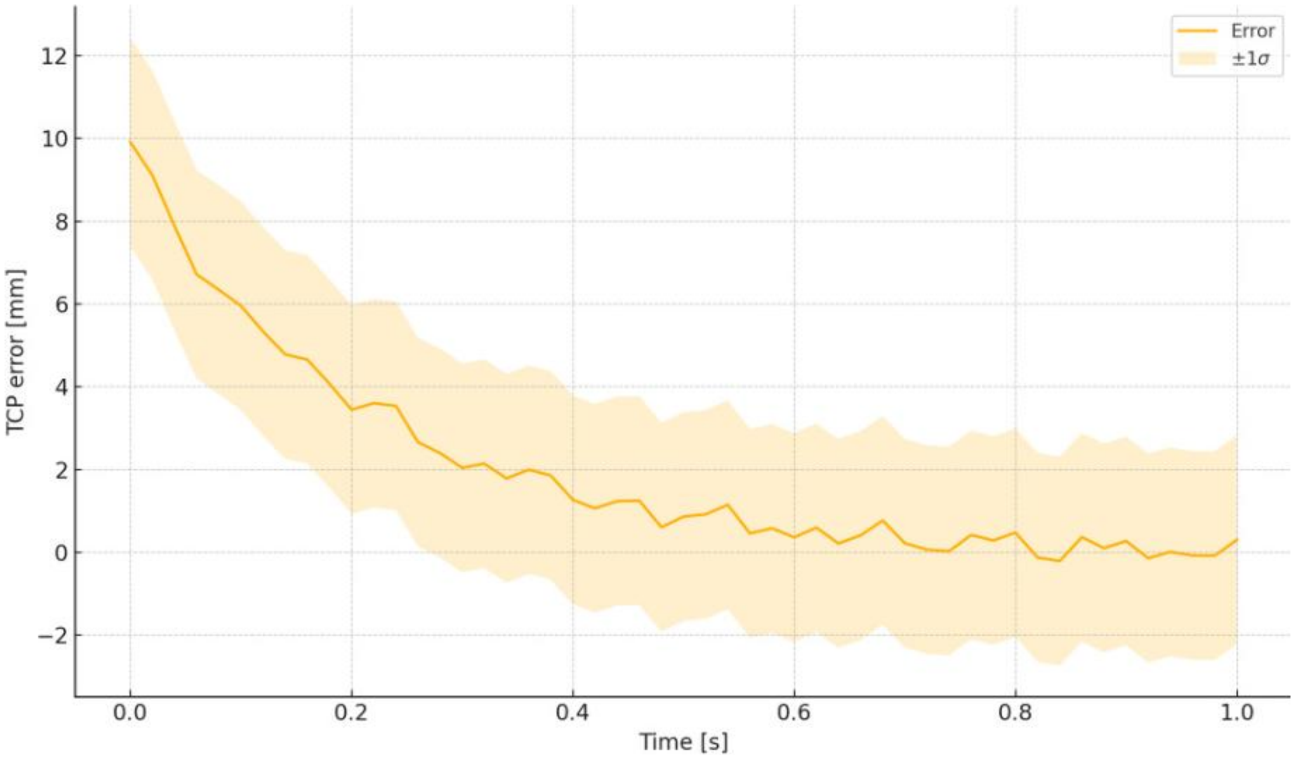


Figure 3 – Error convergence during fine alignment: error drops below 1 mm in ~0.4 s and settles near 0.7 mm

new sensors or control modules to be incorporated with minimal changes.

Compared to a single-camera baseline, the proposed dual-camera + MPC architecture reduced the final positioning error from around **5-6 mm** to **below 1 mm**, while preserving real-time closed-loop operation and adding only a small overhead to the engraving cycle. This demonstrates that integrating multi-source visual feedback with model-based optimal control can deliver substantial gains in precision without sacrificing throughput.

Future work will focus on enhancing adaptability to even less-structured environments. One direction is to integrate learning-based strategies, such as reinforcement learning agents or generative models, to handle cases without fiducial markers or with more unpredictable changes. Another avenue is to extend the system to full 3D trajectories on complex surfaces, potentially incorporating additional sensing like force feedback. We anticipate that the core concepts demonstrated here – multi-source visual feedback combined with optimal control – will remain effective as we scale to more complex scenarios. Ultimately, this work contributes to intelligent industrial automation by showing how vision and model-based control can be synergistically combined to achieve precision and flexibility beyond traditional fixed automation.

Acknowledgements: *This study was funded by the research project BR24992947 – Development of Robots, Scientific, Technical and Software for Flexible Robotization and Industrial Automation (RPA) in Automotive Industrial Enterprises in Kazakhstan Using Artificial Intelligence.*

REFERENCES

1. Wu J., Jin Z., Liu A., Yu L. Vision-based neural predictive tracking control for multi-manipulator systems with parametric uncertainty. ISA Transactions, 2021, vol. 110, pp. 247-257.
2. Gabler V., Huber G., Wollherr D. A force-sensitive grasping controller using tactile gripper fingers and an industrial position-controlled robot. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2022, pp. 770-776.

3. Cao J., et al. Mechanism design and dynamic switching modal control of the wheel-legged separation quadruped robot. Robotica, 2024, vol. 42, no. 3, pp. 660-683.
4. Li H., et al. An ultrasound-guided miniature hand-eye integrated robot for percutaneous needle placement. Journal of Mechanisms and Robotics, 2025, vol. 17, no. 6, p. 065001.
5. Eichenbaum J., et al. Towards a lifelong mapping approach using Lanelet2 for autonomous open-pit mine operations. In: Proceedings of IEEE CASE, 2023, pp. 1-8.
6. Si J., et al. A UAV autonomous landing system integrating locating, tracking, and landing in the wild environment. Journal of Intelligent & Robotic Systems, 2024, vol. 110, no. 2, p. 51.
7. Yan J., Liu M., Jin L. Cerebellum-inspired model predictive control for redundant manipulators with unknown structure information. IEEE Transactions on Cognitive and Developmental Systems, 2023, vol. 16, no. 3, pp. 1198-1210.

ROS 2 негізінде көру сервобасқаруы мен модельдік болжамды басқаруды біріктіретін роботтандырылған таңбалау үшін модульдік басқару архитектурасы

[1,2]**ТАҢАТОВ Айбар Ілиясұлы,** *докторант, jazbalar@gmail.com,*
[1,3]**ОМАРОВ Батырхан Султанович,** *PhD, профессор, batyahan@gmail.com,*
[1,4]***КУАТОВА Молдир Жангелдиевна,** *PhD, профессор, kuatova.moldyr@gmail.com,*
[5]**МЕНДЕШ Матеуш,** *PhD, адъюнкт-профессор, mmendes@isr.uc.pt,*
[1]**АБЕКОВА Айдана Жумагалиевна,** *докторант, Abekova310792@gmail.com,*
[1]Академик Ө.А. Жолдасбеков атындағы механика және машинатану институты, Шевченко көшесі, 29, Алматы, Қазақстан,
[2]«Әл-Фараби атындағы Қазақ ұлттық университеті» КеАҚ, әл-Фараби даңғылы, 71, Алматы, Қазақстан,
[3]«Халықаралық ақпараттық технологиялар университеті» АҚ, Манас көшесі, 34/1, Алматы, Қазақстан,
[4]«Халықаралық инженерлік-технологиялық университеті» ЖШС, әл-Фараби даңғылы, 89/21, Алматы, Қазақстан,
[5]Коимбра политехникалық университеті, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, Коимбра, Португалия,
**автор-корреспондент.*

Аңдатпа. *Бұл жұмыста жоғары дәлдікті автономды таңбалау үшін ROS 2 негізіндегі интеллектуалды басқару архитектурасы ұсынылады. Жүйе субмиллиметрлік дәлдікке жету мақсатында модельдік предиктивті басқаруды (MPC) екі камералы көру жүйесімен біріктіреді. Көпдеңгейлі басқару стегі үстіңгі камерадан алынған ғаламдық көруді және құралға орнатылған камерадан жергілікті түзетуді үйлестіріп, маркерлер мен «көз-қол» калибрлеуі арқылы алты еркіндік дәрежесінде дәл позицияны анықтайды. Траектория жоспарлау CasADi оптимизациялық пакеті арқылы жұмыс аймағы мен жетек шектеулерін ескеріп, нақты уақытта бейімделген қозғалыс генерациясын қамтамасыз етеді. Архитектура RS-485 желісіндегі STM32 контроллерлері бар 6 DOF коботта іске асырылып, 0,7 мм-ге дейінгі қателік пен шамамен 8 Гц жабық цикл жиілігіне қол жеткізілді. Нәтижелер көру мен оптималды басқаруды ROS 2 модульдік ортада біріктірудің өнеркәсіптік автоматизацияда тиімділігін дәлелдейді.*

Кілт сөздер: *ROS 2, көру арқылы басқару, модельдік болжамды басқару, роботтандырылған таңбалау, қос камералы кері байланыс, STM32 микроконтроллері, траекторияны оңтайландыру, «қол-көз» калибрлеуі, нақты уақыттағы басқару, жоғары дәлдікті робототехника.*

Модульная архитектура управления на базе ROS 2 с интеграцией визуального сервопривода и моделирующего предиктивного управления для роботизированной маркировки

[1,2]**ТАҢАТОВ Айбар Ілиясұлы,** *докторант, jazbalar@gmail.com,*

[1,3]**ОМАРОВ Батырхан Султанович,** *PhD, профессор, batyahan@gmail.com,*
[1,4]***КУАТОВА Молдир Жангелдиевна,** *PhD, профессор, kuatova.moldyr@gmail.com,*
[5]**МЕНДЕШ Матеуш,** *PhD, адъюнкт-профессор, mmendes@isr.uc.pt,*
[1]**АБЕКОВА Айдана Жумагалиевна,** *докторант, Abekova310792@gmail.com,*
[1]Институт механики и машиноведения имени академика У.А. Джолдасбекова, ул. Шевченко, 29, Алматы, Казахстан,
[2]НАО «Казахский национальный университет имени аль-Фараби», пр. аль-Фараби, 71, Алматы, Казахстан,
[3]АО «Международный университет информационных технологий», ул. Манаса, 34/1, Алматы, Казахстан,
[4]ТОО «Международный инженерно-технологический университет», пр. аль-Фараби, 89/21, Алматы, Казахстан,
[5]Политехнический университет Коимбры, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, Коимбра, Португалия,

**автор-корреспондент.*

Аннотация. *В работе представлена интеллектуальная архитектура управления на базе ROS 2 для высокоточных автономных операций маркировки. Система интегрирует предиктивное управление (MPC) и двухкамерное визуальное сопровождение для достижения субмиллиметровой точности. Многоуровневый стек управления сочетает глобальное восприятие с верхней камеры и локальную коррекцию с камеры на инструменте, обеспечивая точное определение положения в шести степенях свободы через маркеры и калибровку «глаз-рука». Планирование траектории формулируется как задача оптимизации в CasADi с учётом ограничений рабочего пространства и приводов, что позволяет в реальном времени генерировать адаптивные движения. Архитектура реализована на 6-осевом коботе с контроллерами STM32 (сеть RS-485), достигая ошибки до 0,7 мм и частоты замкнутого контура около 8 Гц. Результаты подтверждают эффективность подхода и демонстрируют потенциал сочетания зрения и оптимального управления в модульной среде ROS 2 для промышленной автоматизации.*

Ключевые слова: *ROS 2, визуальное слежение, моделирующее предиктивное управление, роботизированная маркировка, двухкамерная обратная связь, микроконтроллер STM32, оптимизация траектории, калибровка «рука-глаз», управление в реальном времени, высокоточная робототехника.*

REFERENCES

1. Wu J., Jin Z., Liu A., Yu L. Vision-based neural predictive tracking control for multi-manipulator systems with parametric uncertainty. ISA Transactions, 2021, vol. 110, pp. 247-257.
2. Gabler V., Huber G., Wollherr D. A force-sensitive grasping controller using tactile gripper fingers and an industrial position-controlled robot. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2022, pp. 770-776.
3. Cao J., et al. Mechanism design and dynamic switching modal control of the wheel-legged separation quadruped robot. Robotica, 2024, vol. 42, no. 3, pp. 660-683.
4. Li H., et al. An ultrasound-guided miniature hand-eye integrated robot for percutaneous needle placement. Journal of Mechanisms and Robotics, 2025, vol. 17, no. 6, p. 065001.
5. Eichenbaum J., et al. Towards a lifelong mapping approach using Lanelet2 for autonomous open-pit mine operations. In: Proceedings of IEEE CASE, 2023, pp. 1-8.
6. Si J., et al. A UAV autonomous landing system integrating locating, tracking, and landing in the wild environment. Journal of Intelligent & Robotic Systems, 2024, vol. 110, no. 2, p. 51.
7. Yan J., Liu M., Jin L. Cerebellum-inspired model predictive control for redundant manipulators with unknown structure information. IEEE Transactions on Cognitive and Developmental Systems, 2023, vol. 16, no. 3, pp. 1198-1210.