A Coordinate Transformation Framework for Vision-Guided Robotic Positioning in Automated Vehicle Body Marking

Olzhas Olzhayev[1], Azhar Tursynova[2], Sayat Ibrayev[3]
Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan[1, 2, 3]
International Information Technology University, Almaty, Kazakhstan[1, 2]

Abstract—**Vision-guided robotic positioning plays a critical role in modern manufacturing by enabling robots to accurately adapt to variations in workpiece position. This study proposes a coordinate transformation framework for automated vehicle body marking that converts visually detected target coordinates into executable robot motions through a unified processing pipeline. The proposed framework integrates visual target localization, homogeneous coordinate transformation, Tool Center Point (TCP) pose generation, and inverse kinematics to achieve accurate robotic positioning. Experimental validation was performed using a vision-guided robotic marking platform consisting of an AR4-MK3 six-degree-of-freedom manipulator, an overhead monocular camera, and a 50 W electromagnetic marking device. Five target positions with two repeated trials per position were used to evaluate positioning accuracy, repeatability, and computational performance. The proposed framework achieved an average positioning error of 2.19 mm, a maximum error of 2.69 mm, and a total processing time of approximately 27 ms, demonstrating stable and reliable robotic positioning during the marking process. The proposed framework provides a practical and computationally efficient solution for transforming visual information into robot-executable poses and offers a promising foundation for future vision-guided robotic marking systems operating in industrial environments.**

Keywords*—*Coordinate transformation; vision-guided robotics; robotic positioning; vehicle body marking; monocular vision; tool center point (TCP); industrial robotics

I. Introduction

The rapid evolution of Industry 4.0 has significantly accelerated the adoption of intelligent robotic systems in modern manufacturing, where high precision, flexibility, and production efficiency have become essential requirements. In the automotive industry, industrial manipulators are extensively employed in welding, assembly, painting, inspection, machining, and marking operations, enabling manufacturers to improve product quality while reducing production time and operational costs [1]. Among these manufacturing processes, automated vehicle body marking represents a critical operation because permanent identification information must be applied at predefined locations with high positional accuracy and repeatability. Even minor positioning errors may result in defective markings, reduced readability, or costly rework during production.

Traditional robotic marking systems generally execute trajectories generated through offline programming, assuming that every workpiece occupies an identical position within the robot workspace. However, practical manufacturing environments inevitably introduce uncertainties caused by fixture tolerances, conveyor deviations, thermal deformation, mechanical vibration, and accumulated assembly errors [2]. These positional variations reduce the effectiveness of predefined robot programs and frequently require manual teaching, trajectory modification, or repeated calibration, thereby limiting production flexibility and increasing manufacturing costs.

To improve adaptability, vision-guided robotic systems have become an important research direction in intelligent manufacturing. By integrating industrial cameras with robotic manipulators, visual sensing enables robots to estimate the actual position of workpieces and dynamically adjust robot motion according to the current manufacturing environment [3]. Vision-guided positioning has demonstrated considerable advantages in robotic assembly, machining, inspection, and manipulation because it reduces dependence on precisely fixed fixtures while improving positioning accuracy under varying operating conditions [4]. Nevertheless, the successful implementation of such systems depends not only on reliable visual perception but also on the accurate transformation of visual information into robot-executable motions.

The transformation of spatial information between multiple coordinate systems constitutes one of the fundamental problems in vision-guided robotics. During an automated marking operation, the target location detected by the vision system must be sequentially represented in the camera coordinate system, transformed into the robot base coordinate system, converted to the end-effector coordinate frame, and finally expressed as an executable Tool Center Point (TCP) pose. Errors introduced at any stage of this transformation pipeline directly affect the final positioning accuracy of the marking tool and consequently influence the quality of the manufacturing process.

Numerous studies have investigated individual components of this transformation process. Classical hand-eye calibration algorithms established reliable mathematical relationships between camera and robot coordinate systems [5–7], while subsequent research improved calibration accuracy through optimization techniques [8], comparative evaluations [9], and learning-based three-dimensional vision approaches [10]. These studies have significantly advanced camera–robot calibration; however, they primarily investigate calibration as an isolated problem rather than considering the complete sequence of transformations required for practical robotic positioning.

In industrial vehicle body marking, accurate robotic positioning requires considerably more than camera calibration alone. The entire processing chain – including visual target localization, coordinate transformation, pose generation, and robot motion execution – must operate consistently to ensure that the marking tool reaches the desired location despite workpiece displacement and positioning uncertainties. Although these individual components have been widely studied, relatively few works integrate them into a unified coordinate transformation framework specifically designed for automated vehicle body marking.

To address this limitation, this study proposes a coordinate transformation framework for vision-guided robotic positioning in automated vehicle body marking. The proposed framework integrates visual target localization, homogeneous coordinate transformation, robot pose generation, and robotic positioning into a unified processing pipeline that transforms a visually detected marking target into an executable robot pose. Unlike conventional approaches that primarily evaluate calibration accuracy, the proposed framework emphasizes the complete transformation process required for practical robotic marking applications and experimentally validates its positioning performance under different workpiece displacement conditions.

II. Related Works

Vision-guided robotic positioning has become an important component of flexible manufacturing systems because visual feedback enables manipulators to respond to the actual position and orientation of objects in the workspace. Hutchinson et al. established the fundamental principles of visual servoing and classified vision-based robot control approaches according to the manner in which visual information is incorporated into the control loop [11]. Chaumette and Hutchinson subsequently provided a systematic description of image-based and position-based visual servoing, emphasizing that the selection of visual features and coordinate representations directly affects system stability, convergence, and positioning accuracy [12]. These studies formed the theoretical basis for modern camera-guided robotic systems; however, their primary focus was general visual control rather than application-specific coordinate transformation for industrial marking operations.

The accuracy of a vision-guided system strongly depends on the geometric relationship between the camera, robot base, end effector, and workpiece coordinate frames. Homogeneous transformation matrices are commonly used to represent these relationships and convert visually detected positions into poses executable by a robot controller. Ali et al. investigated simultaneous robot-world and hand-eye calibration and demonstrated that the definition of coordinate frames and optimization criteria significantly influence the accuracy of the estimated transformations [13]. Enebuse et al. later reviewed a wide range of hand-eye calibration techniques and showed that existing methods differ considerably in their treatment of rotation, translation, measurement noise, and camera configuration [14]. These findings confirm that coordinate transformation cannot be treated as a purely mathematical operation because errors introduced at the calibration stage propagate to the final robot pose.

Experimental comparisons have therefore received considerable attention in recent calibration research. Enebuse et al. evaluated several commonly used hand-eye calibration algorithms under controlled conditions and demonstrated that their accuracy depends on the robot motion set, calibration target observations, and error evaluation method [15]. Strobl and Hirzinger proposed an optimal hand-eye calibration approach that explicitly considers the uncertainty of robot and camera measurements [16]. More recently, learning-based three-dimensional vision has been employed to automate calibration target detection and reduce manual intervention during camera–robot calibration [17]. Although these studies improve the estimation of transformation parameters, they predominantly investigate calibration as an independent procedure rather than as part of a complete pipeline that includes target localization, pose generation, trajectory execution, and validation of the final tool position.

Application-oriented studies have shown that calibration and coordinate transformation become particularly critical in robotic processes requiring accurate contact or alignment with a workpiece. Zhan and Wang developed a hand-eye calibration and positioning method for robotic drilling, in which visual measurements were transformed into robot coordinates to maintain the required tool orientation and distance from an aircraft panel [18]. Li et al. proposed a visually guided error-compensation method for robotic machining and demonstrated that visual measurement can improve the pose accuracy of industrial manipulators during drilling and milling operations [19]. Similar challenges arise in vehicle body marking because the marking tool must be aligned with a predefined surface location while compensating for workpiece displacement, robot positioning errors, and uncertainties in the camera-to-robot transformation.

Despite the progress achieved in visual servoing, hand-eye calibration, and vision-guided machining, the reviewed studies mainly address visual control, calibration parameter estimation, or positioning-error compensation as separate problems. Limited attention has been given to unified frameworks that transform a visually detected target through the camera, robot-base, flange, and tool coordinate systems and then generate an executable pose for an automated marking process. Furthermore, most calibration studies evaluate transformation residuals or reprojection errors, whereas practical marking systems require validation of the final end-effector position under different workpiece displacements. This research gap motivates the proposed coordinate transformation framework, which integrates visual target localization, homogeneous coordinate transformation, target-pose generation, and robotic positioning for automated vehicle body marking. The need for such integrated approaches is also supported by recent industrial vision studies demonstrating that calibration accuracy, target localization, and robot motion execution must be evaluated as interconnected stages rather than isolated components [20].

III. Materials and Methods

A. Vision-Guided Robotic Marking System

The experimental vision-guided robotic marking system developed in this study is shown in Fig. 1. The platform consists of an AR4-MK3 six-degree-of-freedom robotic manipulator, an electromagnetic marking device mounted at the robot flange, an overhead monocular camera, and a computer-based control unit. The objective of the system is to validate the proposed coordinate transformation framework for accurate robotic positioning during automated vehicle body marking. The monocular camera continuously observes the workspace and captures images of the workpiece. After detecting the predefined marking target, the corresponding image coordinates are processed by the proposed framework, which transforms the target position from the camera coordinate system into the robot coordinate system. Based on the transformed coordinates, the corresponding Tool Center Point (TCP) pose is generated and transmitted to the robot controller for motion execution.

Fig. 1. Vision-guided robot equipped with an electromagnetic marking device.

Unlike conventional robotic marking systems based on offline programming, the proposed platform determines the target position online using visual information and automatically updates the robot pose before executing the marking operation. This enables the manipulator to compensate for workpiece displacement while preserving the required positioning accuracy. The overall processing pipeline consists of four consecutive stages: target localization, coordinate transformation, robot pose generation, and robotic positioning.

B. Coordinate Systems

To accurately transform the visually detected target into an executable robot pose, the proposed framework defines a series of coordinate systems that describe the spatial relationship between the vision sensor, robotic manipulator, and marking tool. These coordinate systems establish a common geometric representation that enables consistent transformation of target positions throughout the robotic positioning process.

As illustrated in Fig. 2, four coordinate systems are employed in the proposed framework. The camera coordinate system {C} is attached to the optical center of the monocular camera and is used to represent the three-dimensional position of the detected marking target after image processing. The robot base coordinate system {B} is fixed at the robot base and serves as the global reference frame for robot motion planning and positioning.

The flange coordinate system {F} is attached to the robot flange and describes the pose of the robot wrist during manipulator motion. Finally, the tool center point (TCP) coordinate system {T} is defined at the tip of the electromagnetic marking device and represents the final position and orientation required to perform the marking operation. Since the marking device is rigidly mounted to the robot flange, the transformation between the flange and TCP coordinate systems remains constant throughout the experiments.

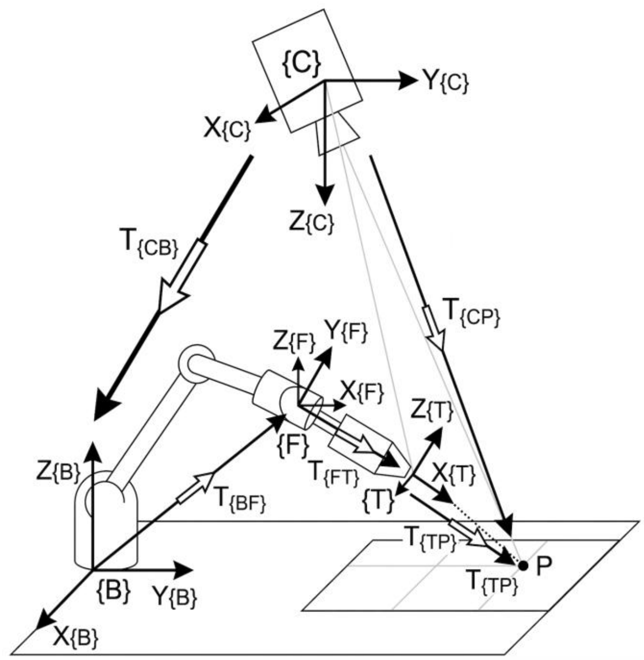

Fig. 2. Coordinate systems used in the proposed framework.

The objective of the proposed framework is to determine the TCP pose corresponding to the visually detected marking target by sequentially transforming the target coordinates from the camera reference frame to the robot base frame and subsequently to the tool coordinate system. This transformation process forms the foundation of the mathematical model presented in the following section.

C. Coordinate Transformation Framework

The proposed coordinate transformation framework is designed to convert a visually detected marking target into an executable robot pose through a sequence of geometric transformations between multiple coordinate systems. Instead of relying on predefined robot trajectories, the framework dynamically computes the target pose from visual observations, enabling automatic compensation for workpiece displacement and positioning uncertainties.

The framework consists of four sequential stages, which are demonstrated in Fig. 3. First, the vision system acquires an image of the workspace and determines the location of the predefined marking target. The detected target is initially represented in the camera coordinate system. Second, the target coordinates are transformed into the robot base coordinate system using the geometric relationship established during system calibration. Third, the transformed target position is converted into the corresponding Tool Center Point (TCP) pose by considering the fixed spatial relationship between the robot flange and the mounted marking device. Finally, the generated TCP pose is transmitted to the robot controller, which computes the robot joint configuration and executes the positioning motion.

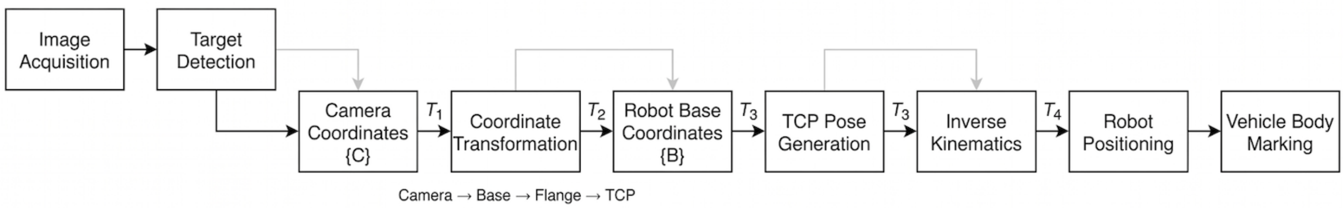

Fig. 3. Overall coordinate transformation framework for vision-guided robotic positioning.

The proposed framework considers the complete transformation process from visual target detection to robotic positioning. By integrating target localization, coordinate transformation, pose generation, and robot motion into a unified processing pipeline, the framework provides a consistent geometric representation throughout the entire marking procedure. This unified formulation simplifies the implementation of vision-guided robotic positioning while improving the adaptability of the system to variations in workpiece position.

D. Mathematical Formulation

The proposed framework computes the robot Tool Center Point (TCP) pose by sequentially transforming the detected target coordinates from the camera coordinate system to the robot base coordinate system. The transformations are represented using homogeneous transformation matrices.

The detected target is represented in homogeneous coordinates as:

$$\mathrm{P_C} = [\mathrm{x_C}; \mathrm{y_C}; \mathrm{z_C}; 1] \tag{1}$$

where, P_C is the target position in the camera coordinate system.

The target coordinates are transformed into the robot base coordinate system according to:

$$\mathrm{P_B} = \mathrm{T}_{BC}\mathrm{P_C} \tag{2}$$

where, T_{BC} denotes the homogeneous transformation matrix from the camera frame to the robot base frame.

The transformation matrix is defined as:

$$\mathrm{T_{BC}} = \begin{bmatrix} R_{BC} & t_{BC} \\ 0 & 1 \end{bmatrix} \tag{3}$$

where, R_{BC} is the rotation matrix and t_{BC} is the translation vector.

The TCP pose is obtained by applying the fixed tool transformation:

$$\mathrm{T_{TCP}} = \mathrm{T_{BF}}\mathrm{T_{FT}} \tag{4}$$

where, T_{BF} is the flange pose with respect to the robot base and T_{FT} is the constant transformation from the flange to the TCP.

The desired robot pose is calculated as:

$$T_{Target} = T_{BC}T_{CT}T_{TT} \tag{5}$$

where, T_{CT} transforms the detected target from the camera frame and T_{TT} represents the tool offset.

The robot joint configuration is obtained as:

$$\mathrm{q} = \mathrm{IK}(\mathrm{T_{Target}}) \tag{6}$$

where, q denotes the robot joint vector, and *IK* is inverse kinematics.

E. Experimental Procedure

The experimental validation was conducted using the developed vision-guided robotic marking platform. The system consisted of an AR4-MK3 six-degree-of-freedom robotic manipulator (maximum payload 1.9 kg and repeatability up to 0.2 mm), an overhead monocular camera, and a 50 W electromagnetic marking device mounted on the robot end-effector.

During each experiment, the camera detected the target position, after which the proposed coordinate transformation framework computed the corresponding TCP pose and generated the robot motion command. The robot then positioned the marking tool at the desired location. The framework was evaluated in terms of positioning accuracy, repeatability, and computation time under multiple target locations distributed throughout the workspace.

IV. RESULTS

The proposed coordinate transformation framework was experimentally validated using the developed vision-guided robotic marking platform. The experimental setup consisted of an AR4-MK3 six-degree-of-freedom robotic manipulator, an Intel® RealSense™ D455 RGB-D camera mounted above the workspace for visual target detection, a 50 W electromagnetic marking device attached to the robot end-effector, and a computer-based control unit responsible for image processing, coordinate transformation, and robot motion planning. During the experiments, the RealSense D455 continuously monitored the working area and acquired visual information required for target localization. The detected target coordinates were subsequently transformed into the robot coordinate system using the proposed framework, enabling automatic generation of the Tool Center Point (TCP) pose and execution of the marking operation. The complete experimental platform is presented in Fig. 4.

Fig. 4. Experimental vision-guided robotic marking platform.

During the experiments, the Intel® RealSense™ D455 camera continuously monitored the experimental workspace and acquired visual data of the vehicle body surface. The captured RGB images were processed to identify the predefined marking target and determine its position within the camera coordinate system. The extracted target coordinates were then supplied to the proposed coordinate transformation framework, which sequentially transformed them into the robot base coordinate system, generated the corresponding Tool Center Point (TCP) pose, and transmitted the resulting robot motion command to the manipulator controller. This automated processing pipeline enabled accurate positioning of the marking tool while compensating for variations in the workpiece location. An example of the camera view used for target localization during the experiments is presented in Fig. 5.

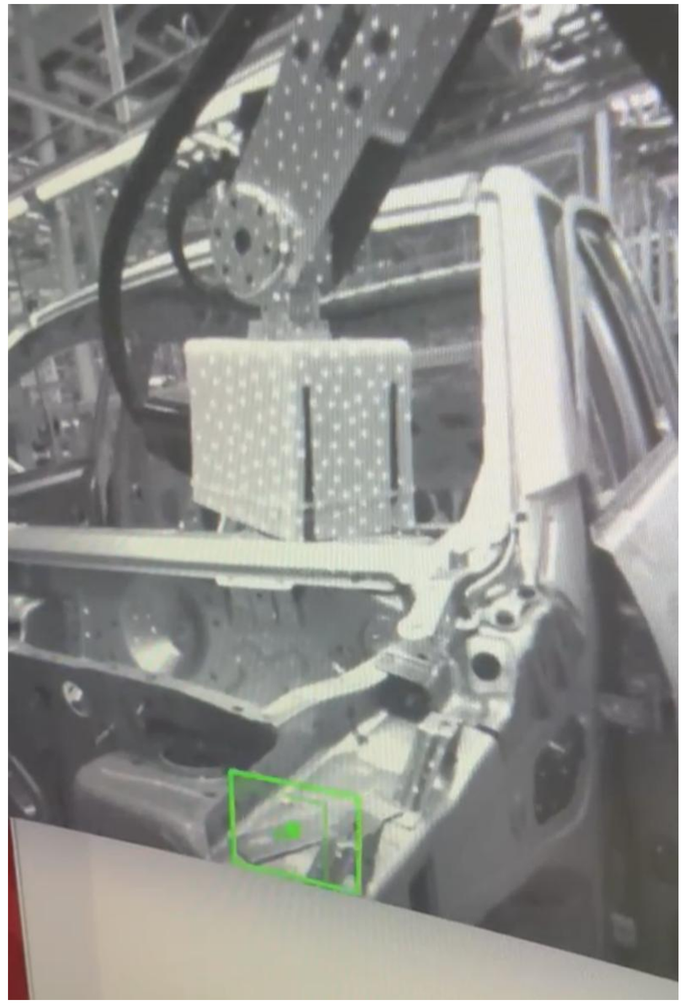

Fig. 5. Camera view of the experimental workspace used for target localization.

The computed robot pose was subsequently executed by the AR4-MK3 manipulator, positioning the electromagnetic marking device at the desired location on the workpiece. Upon reaching the target pose, the marking operation was automatically performed, producing a permanent mark at the predefined position. The experimental marking result is presented in Fig. 6, demonstrating the successful implementation of the complete vision-guided robotic positioning pipeline, from visual target detection and coordinate transformation to accurate robot positioning and automated marking.

Fig. 6. Experimental marking result obtained using the proposed framework.

The positioning accuracy was evaluated at multiple target locations distributed throughout the robot workspace. The deviations between the desired and actual marking positions along the X, Y, and Z axes are summarized in Table I, while the overall experimental performance, including repeatability and computational efficiency, is presented in Table II.

TABLE I. POSITIONING ERRORS OF THE PROPOSED VISION-GUIDED ROBOTIC POSITIONING FRAMEWORK AT DIFFERENT TARGET POINTS

Target point	X error (mm)	Y error (mm)	Z error (mm)	Total error (mm)
P1	0.8	1.2	0.9	1.71
P2	1.3	0.9	1.1	1.93
P3	1.6	1.4	0.8	2.29
P4	1.1	1.8	1.0	2.34
P5	1.9	1.5	1.2	2.69
Mean	1.34	1.36	1.0	2.19
Standard deviation	0.42	0.35	0.16	0.39
Maximum	1.90	1.80	1.20	2.69

The results demonstrated that the proposed framework consistently transformed the visually detected target coordinates into executable robot poses. The measured errors remained within the acceptable range for experimental vehicle body marking applications. Repeated trials at the same target positions also showed stable robot positioning and consistent marking performance.

In addition to positioning accuracy, the framework was evaluated in terms of repeatability and computation time. The computation time included target localization, coordinate transformation, and TCP pose generation. The summarized performance indicators are presented in Table II.

TABLE II. OVERALL EXPERIMENTAL PERFORMANCE OF THE PROPOSED VISION-GUIDED ROBOTIC POSITIONING FRAMEWORK

Performance indicator	Value
Number of target positions	5
Trials per target	2
Mean positioning error (mm)	2.19
Maximum positioning error (mm)	2.69
Repeatability / standard deviation (mm)	0.39
Target detection time (ms)	25
Coordinate transformation time (ms)	1.3
Total processing time (ms)	27

Overall, the experimental results confirmed that the proposed coordinate transformation framework provides accurate and repeatable conversion of visual target coordinates into robot motion commands. The successful execution of the marking operation also demonstrated the practical feasibility of the developed system for automated positioning and marking tasks.

V. DISCUSSION

The experimental results demonstrate that the proposed coordinate transformation framework provides a reliable and computationally efficient solution for vision-guided robotic positioning in automated vehicle body marking. By integrating visual target localization, homogeneous coordinate transformation, TCP pose generation, inverse kinematics, and robot motion execution within a unified workflow, the framework successfully converts visually detected target coordinates into executable robotic motions. This integrated formulation is particularly important in flexible manufacturing environments, where variations in workpiece position can reduce the effectiveness of conventional preprogrammed robotic trajectories [21].

The mean positioning error of **2.19 mm**, with a maximum error of **2.69 mm**, indicates that the proposed framework can maintain consistent positioning performance across different locations within the evaluated workspace. Recent developments in vision-guided industrial robotics similarly emphasize that accurate spatial transformation between sensing and robot coordinate systems is fundamental to reliable manipulation and manufacturing operations [22]. The standard deviation of **0.39 mm** further indicates relatively stable positioning behavior across the evaluated targets. Repeatable positioning is important in automated manufacturing because errors originating from perception, calibration, coordinate transformation, and manipulator motion can propagate to the final tool location [23].

From an algorithmic perspective, the proposed framework can be viewed as a sequential perception–transformation–planning pipeline. The target-localization algorithm first extracts the required spatial information from the camera observation, after which homogeneous transformation algorithms map the detected coordinates from the camera frame to the robot base frame. Inverse kinematics then determines a feasible joint configuration corresponding to the desired TCP pose. Such modular processing is advantageous because individual algorithms can be independently improved without redesigning the complete robotic system [24].

The obtained positioning performance represents the cumulative accuracy of this complete vision-to-motion pipeline rather than camera calibration accuracy alone. Camera–robot calibration establishes the geometric relationship between coordinate systems, whereas final positioning additionally depends on visual localization, transformation parameters, TCP calibration, manipulator kinematics, and mechanical accuracy [25]. Consequently, evaluating only reprojection or calibration residuals may not adequately characterize the practical performance of a robotic marking system. The proposed experiments instead evaluate the final positioning outcome after the target coordinates have passed through the complete computational and robotic execution chain.

Homogeneous transformation matrices provide the mathematical foundation for the coordinate-conversion algorithm because rotations and translations between the camera, robot base, flange, and TCP frames can be represented consistently within the same formulation [26]. Accurate estimation of these relationships is especially important when visual measurements must be converted into physically executable robot poses. The proposed sequential transformation algorithm consequently provides a relatively transparent approach in which each geometric operation can be inspected, calibrated, and validated independently.

Machine learning represents an important opportunity for extending this framework beyond its current primarily geometric implementation. In the present system, the coordinate transformation stage is deterministic and model-based; however, machine-learning algorithms can complement this stage by improving visual target detection, feature localization, pose estimation, calibration refinement, and positioning-error compensation [27-29]. Learning-based three-dimensional vision has already demonstrated potential for automating camera–robot calibration and reducing dependence on manually defined calibration procedures [30]. Incorporating such algorithms into the perception component could improve robustness when targets are observed under variations in illumination, viewpoint, surface appearance, or partial occlusion [31].

Deep learning algorithms are particularly relevant to the visual localization stage. Convolutional neural networks, object detectors, segmentation networks, and vision-transformer-based models could be employed to identify marking regions or reference features directly from camera images [32]. Instead of depending exclusively on predefined visual features, a trained model could learn discriminative representations from annotated examples and estimate target positions under more complex manufacturing conditions [33]. Nevertheless, introducing machine learning would require representative training data covering realistic variations in vehicle surfaces, camera viewpoints, illumination, occlusion, and target appearance to avoid degradation under domain shift.

Machine learning could also be applied after target detection to compensate for systematic positioning errors. A regression model could learn a correction function from the desired target position, camera observation, robot configuration, and measured positioning residuals [34]. The predicted correction could then be applied to the geometrically calculated TCP pose before robot execution. Such a hybrid strategy would preserve the interpretability of the homogeneous coordinate transformation while allowing data-driven algorithms to compensate for nonlinear errors that are difficult to represent explicitly using conventional calibration equations [35]. This combination of analytical geometry and machine learning may therefore provide a useful direction for improving positioning accuracy without replacing the underlying coordinate framework.

The computational results are also encouraging for future machine-learning integration. The complete pipeline required approximately **27 ms**, with target detection accounting for approximately **25 ms** and coordinate transformation requiring only **1.3 ms**. The small computational cost of the geometric transformation leaves a substantial portion of the processing architecture available for more advanced perception algorithms. Efficient machine-learning models could potentially be incorporated while maintaining near-real-time operation, particularly when lightweight neural architectures, model compression, or hardware acceleration are employed.

Algorithm selection should nevertheless consider the trade-off between accuracy, computational complexity, interpretability, and data requirements. Classical image-processing algorithms can provide predictable behavior when the target appearance and illumination conditions are highly controlled, whereas machine-learning algorithms generally provide greater adaptability when visual conditions are more heterogeneous. A hybrid architecture in which machine learning performs robust target localization and deterministic geometric algorithms perform coordinate transformation may therefore offer a practical compromise for industrial deployment [36]. This design also preserves explicit geometric relationships among the coordinate frames, which facilitates calibration verification and troubleshooting.

The proposed framework differs from approaches that consider hand–eye calibration as an isolated optimization problem. Recent calibration algorithms have achieved improvements through robust optimization, uncertainty modeling, automated target detection, and learning-based estimation [33]. However, accurate calibration alone does not guarantee accurate industrial operation because the resulting transformation parameters must subsequently interact with TCP geometry, inverse kinematics, robot control, and mechanical characteristics. The present framework therefore emphasizes the complete sequence from camera observation to final positioning and evaluates its performance according to the actual marking-tool location.

The experimental marking results further demonstrate the practical feasibility of combining visual sensing, coordinate-transformation algorithms, and robotic motion control. Vision-guided systems improve manufacturing flexibility by allowing manipulators to respond to observed workpiece positions instead of depending entirely on mechanically constrained fixtures [34]. In future implementations, machine-learning-based visual perception could further increase this adaptability by recognizing multiple target types and estimating their positions under more variable production conditions.

Several limitations should nevertheless be acknowledged. The experimental evaluation included only **five target positions with two trials per target**, which demonstrates feasibility but is insufficient for comprehensive statistical characterization across the complete robot workspace. Transformation accuracy can vary according to the spatial distribution of calibration and evaluation poses [35]. Future experiments should therefore include substantially more target locations, repeated trials, different TCP orientations, and positions near the limits of the robot workspace.

The experiments were also conducted under controlled conditions with a stationary workpiece. Industrial vehicle body marking may involve illumination changes, reflections from metallic surfaces, vibration, partial occlusion, camera displacement, and moving components. These conditions represent an important motivation for incorporating machine-learning and deep-learning algorithms into future versions of the framework because data-driven perception can potentially provide greater robustness than manually specified visual features. However, machine-learning performance must be validated using independent test conditions to ensure that improvements are not restricted to the environments represented in the training data.

Future development should consequently investigate a **hybrid geometric and machine-learning framework** rather than replacing the coordinate-transformation model entirely. Machine learning can be used for target detection, feature extraction, pose estimation, calibration refinement, and residual-error prediction, while homogeneous transformations and inverse-kinematics algorithms can retain responsibility for physically interpretable coordinate conversion and motion generation. Adaptive calibration and online learning could additionally compensate for gradual changes caused by camera displacement, tool replacement, mechanical wear, or environmental variation [36]. Reinforcement-learning algorithms could also be investigated for adaptive trajectory generation and motion optimization when the robot must operate around obstacles or dynamically changing workpieces.

Overall, the experimental findings indicate that the proposed framework provides a practical foundation for converting visual target information into robot-executable positioning commands. The combination of a **2.19 mm mean positioning error**, **0.39 mm standard deviation**, and approximately **27 ms total processing time** demonstrates a favorable balance between positioning accuracy, stability, and computational efficiency. The present deterministic coordinate-transformation algorithms provide an interpretable geometric foundation, while the future integration of machine learning, deep visual models, adaptive calibration, and data-driven error-compensation algorithms could substantially improve robustness and adaptability. This hybrid direction offers a promising pathway toward intelligent vision-guided robotic marking systems capable of operating under the variable conditions encountered in automotive manufacturing.

VI. Conclusion

This study presented a coordinate transformation framework for vision-guided robotic positioning in automated vehicle body marking. The proposed framework integrates visual target localization, coordinate transformation, TCP pose generation, and inverse kinematics into a unified workflow that converts image-based target coordinates into executable robot motions. Experimental validation using an AR4-MK3 manipulator, an overhead monocular camera, and a 50 W marking device demonstrated reliable positioning with an average error of approximately 2.2 mm, confirming the feasibility of the proposed approach.

Future work will focus on improving the robustness of the framework in dynamic industrial environments through adaptive calibration, continuous target tracking, and multi-camera perception. Additional validation using industrial robotic platforms and larger-scale experiments will further assess the applicability of the proposed framework for automated manufacturing tasks.

Acknowledgment

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of research program “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] Phuyal, S., Bista, D., & Bista, R. (2020). Challenges, opportunities and future directions of smart manufacturing: a state of the art review. Sustainable futures, 2, 100023.

[2] Bathla, G., Bhadane, K., Singh, R. K., Kumar, R., Aluvalu, R., Krishnamurthi, R., ... & Basheer, S. (2022). Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. Mobile information systems, 2022(1), 7632892.

[3] Singh, A., Kalaichelvi, V., & Karthikeyan, R. (2022). A survey on vision guided robotic systems with intelligent control strategies for autonomous tasks. Cogent Engineering, 9(1), 2050020.

[4] Yang, X., Zhou, Z., Sørensen, J. H., Christensen, C. B., Ünalan, M., & Zhang, X. (2023). Automation of SME production with a Cobot system powered by learning-based vision. Robotics and Computer-Integrated Manufacturing, 83, 102564.

[5] Yi, S., Liu, S., Lin, X., Yan, S., Wang, X. V., & Wang, L. (2025). A data-efficient and general-purpose hand–eye calibration method for robotic systems using next best view. Advanced Engineering Informatics, 66, 103432.

[6] Quality, J. O. F. (2024). An Investigation in Analyzing the Food Quality Well-Being for Lung Cancer Using Blockchain through CNN.

[7] Jiang, J., Luo, X., Luo, Q., Qiao, L., & Li, M. (2022). An overview of hand-eye calibration. The International Journal of Advanced Manufacturing Technology, 119(1), 77-97.

[8] Zhu, D., Wu, H., Ding, T., & Hua, L. (2024). Point cloud registration-enabled globally optimal hand–eye calibration. IEEE/ASME Transactions on Mechatronics, 30(4), 2586-2597.

[9] Tan, G., Du, J., Guo, R., Wang, W., & Ju, Z. (2025). Robust hand-eye calibration with a single plane for 3D robot measurement. IEEE Transactions on Instrumentation and Measurement.

[10] Li, L., Yang, X., Wang, R., & Zhang, X. (2024). Automatic robot hand-eye calibration enabled by learning-based 3d vision. Journal of Intelligent & Robotic Systems, 110(3), 130.

[11] Hutchinson, S., Hager, G. D., & Corke, P. I. (1996). A tutorial on visual servo control. IEEE transactions on robotics and automation, 12(5), 651-670.

[12] Chaumette, F., & Hutchinson, S. (2006). Visual servo control. I. Basic approaches. IEEE robotics & automation magazine, 13(4), 82-90.

[13] Ali, I., Suominen, O., Gotchev, A., & Morales, E. R. (2019). Methods for simultaneous robot-world-hand–eye calibration: A comparative study. Sensors, 19(12), 2837.

[14] Enebuse, I., Foo, M., Ibrahim, B. S. K. K., Ahmed, H., Supmak, F., & Eyobu, O. S. (2021). A comparative review of hand-eye calibration techniques for vision guided robots. IEEE Access, 9, 113143-113155.

[15] Enebuse, I., Ibrahim, B. K. K., Foo, M., Matharu, R. S., & Ahmed, H. (2022). Accuracy evaluation of hand-eye calibration techniques for vision-guided robots. Plos one, 17(10), e0273261.

[16] Strobl, K. H., & Hirzinger, G. (2006, October). Optimal hand-eye calibration. In 2006 IEEE/RSJ international conference on intelligent robots and systems (pp. 4647-4653). IEEE.

[17] Ha, J. (2022). Probabilistic framework for hand–eye and robot–world calibration $ AX= YB$. IEEE transactions on robotics, 39(2), 1196-1211.

[18] Zhan, Q., & Wang, X. (2012). Hand–eye calibration and positioning for a robot drilling system. The International Journal of Advanced Manufacturing Technology, 61(5), 691-701.

[19] Li, B., Li, Y., Tian, W., & Liao, W. (2023). Pose accuracy improvement in robotic machining by visually-guided method and experimental investigation. Robotics and Autonomous Systems, 164, 104416.

[20] Ulrich, M., Steger, C., Butsch, F., & Liebe, M. (2024). Vision-guided robot calibration using photogrammetric methods. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 645-662.

[21] Liu, Y., Yi, J., Chi, P., Liao, H., Zhang, Q., & Wang, Z. (2026). Review of trajectory planning for humanoid welding robot manipulators: from fundamentals to industrial applications in intelligent manufacturing. Journal of Intelligent Manufacturing, 1-19.

[22] Slamani, M., Louhichi, B., Arslane, M., & Alawad, M. A. (2026). Smart industrial robots in the transition from automation to human-centric sustainable systems: a comprehensive review of technologies, challenges, and future directions. The International Journal of Advanced Manufacturing Technology, 143(1), 69-96.

[23] Sameh, A., Fanni, M., & Rashad, M. (2025). Advances in intelligent industrial manipulators for smart manufacturing and standardized automation technologies. Discover Robotics, 1(1), 12.

[24] Ward, R., & Sencer, B. (2023). Accurate TCP position and orientation trajectory generation in 6DOF robotic manipulators and CNC machine tools using FIR filtering and Haversine synchronisation. Procedia CIRP, 120, 27-32.

[25] Min, K., Ni, F., Chen, Z., & Liu, H. (2024). A unified calibration method for robot manipulators: hand-eye parameters, kinematic parameters and TCP position calibration. Robotic Intelligence and Automation, 44(6), 897-909.

[26] Yang, K., Zhang, D., Zhao, X., Zhang, Z., Liu, X., Wang, C., ... & Tao, B. (2026). Low-Cost Mirror-Based Self-Calibration with Kinematic Error Compensation for Vision-Guided Composite Mobile Robots. IEEE Transactions on Instrumentation and Measurement.

[27] Omarov, B., Batyrbekov, A., Dalbekova, K., Abdulkarimova, G., Berkimbaeva, S., Kenzhegulova, S., ... & Omarov, B. (2020, December). Electronic stethoscope for heartbeat abnormality detection. In International Conference on Smart Computing and Communication (pp. 248-258). Cham: Springer International Publishing.

[28] Raghunath, M. P., Deshmukh, S., Chaudhari, P., Bangare, S. L., Kasat, K., Awasthy, M., ... & Waghulde, R. R. (2025). PCA and PSO based optimized support vector machine for efficient intrusion detection in internet of things. Measurement: Sensors, 37, 101806.

[29] Omarov, B., Omarov, B., Rakhymzhanov, A., Niyazov, A., Sultan, D., & Baikuvekov, M. (2024). Development of an artificial intelligence-enabled non-invasive digital stethoscope for monitoring the heart condition of athletes in real-time. Retos, 60, 1169-1180.

[30] Kulambayev, B. O., Olzhayev, O. M., Altayeva, A. B., & Zhunisbekova, Z. (2025). A Multi-Scale ROI-Aligned Deep Learning Framework for Automated Road Damage Detection and Severity Assessment. International Journal of Advanced Computer Science & Applications, 16(12), 1117.

[31] Olzhayev, O., Imanbayeva, N., Mamikov, S., & Baibek, B. (2025). A Novel YOLO-Like Multi-Branch Architecture for Accurate Apple Detection and Segmentation Under Orchard Constraints. International Journal of Advanced Computer Science and Applications, 16(11).

[32] Omarov, B., Baikuvekov, M., Sultan, D., Mukazhanov, N., Suleimenova, M., & Zhekambayeva, M. (2024). Ensemble approach combining deep residual networks and BiGRU with attention mechanism for classification of heart arrhythmias. Computers, Materials & Continua, 80(1), 341-359.

[33] Omarov, B., Tursynova, A., & Uzak, M. (2023). Deep learning enhanced internet of medical things to analyze brain computed tomography images of stroke patients. International Journal of Advanced Computer Science and Applications, 14(8).

[34] Choi, Y. W., Lee, J., Lee, Y., Lee, S., Jeong, W., Lim, D. Y., & Lee, S. W. (2025). A vision-guided adaptive and optimized robotic fabric gripping system for garment manufacturing automation. Robotics and Computer-Integrated Manufacturing, 92, 102874.

[35] Gu, Y., Wu, J., & Liu, C. (2025). Error analysis and accuracy evaluation method for coordinate measurement in transformed coordinate system. Measurement, 242, 115860.

[36] Omarov, B. (2025). Deep learning in biomedical image and signal processing: a survey. Computers, Materials & Continua, 85(2), 2195-2253.