

18th CIRP Conference on Intelligent Computation in Manufacturing Engineering  
Rule-based one-shot object tracking for NC-robots under consideration of energy consumption

Rico Löser<sup>a</sup> \*, Leutrim Gjakova<sup>a</sup>, Marco Schumann<sup>a</sup>, Philipp Klimant<sup>a,b</sup>, Martin Dix<sup>a,c</sup>, Yunqi Gu<sup>d</sup>, Ruth Maria Otto<sup>d</sup>

<sup>a</sup>Fraunhofer-Institut für Werkzeugmaschinen und Umformtechnik IWU, Reichenhainer Straße 88, 09126 Chemnitz, Germany

<sup>b</sup>Hochschule Mittweida, Technikum Platz 17, 09648 Mittweida, Germany

<sup>c</sup>Technische Universität Chemnitz, Straße der Nationen 62, 09111 Chemnitz, Germany

<sup>d</sup>Hochschule München, Lothstraße 34, 80335 München, Germany

\* Corresponding author. Tel.: +49-371-5397-1431 ; fax: +49-371-5397-6-1431. E-mail address: Rico.Looser@iwu.fraunhofer.de

Abstract

In order to become climate neutral, significant energy savings must be achieved in European factories while preserving the economic strength. Simultaneously, new technologies are needed to enable the transition from mass production to mass customization. Cognitive robotics is a key component of this proposition to recognize the environment and dynamically adapt manufacturing behavior. The present paper discusses and evaluates a developed position control for tracking operation of NC-robots based on 2D image processing using ECC algorithm. This approach based on one-shot learning in combination with continuous movement of NC-robots allows to reduce cycle time, respectively energy consumption. Finally, the benefits of the proposed solutions are extended to on-the-fly assembly processes.

© 2025 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 18th CIRP Conference on Intelligent Computation in Manufacturing Engineering (CIRP ICME '24)

Keywords: Tracking, ECC (Enhanced Correlation Coefficient), Visual Servoing, Autonomous Robotics, One Shot Learning, On-The-Fly Assembly, 2D Image Processing ; Green, Sustainable & Resource Efficient Manufacturing

1. Motivation and Introduction

The manufacturing industry is facing several challenges. These include - and are not limited to, the change from mass production to mass customization and the ever-increased complexity of production lines. These challenges and their respective solutions are contradictories. On one hand, low batch sizes and complex production lines require flexible production systems and skilled workers. On the other hand, the competition with low-income countries and low defect rates expectation needs autonomous automation to ensure lower cost, autonomous production, and high quality through high repeatability. In the German French research project GreenBotAI, new solutions are being developed that enable industrial robots to adapt automatically to the needs of the production line and master complex and laborious tasks when handling objects. To achieve savings in cycle time and energy consumption, an efficient workflow is required that includes on-the-fly robot applications and data reduction methods. [1]

Nomenclature

$\delta_{ref}(t)$ : 6D reference position (translative position: X, Y, Z;

rotation around X, Y, Z axis respectively: A, B, C)  
 $\delta_{img}(t)$ : 6D position offset (translative position:  $\Delta X, \Delta Y, \Delta Z$ ;  
rotation around X, Y, Z axis respectively:  $\Delta A, \Delta B, \Delta C$ )  
 $\varepsilon(t)$ : target position of image processing  
 $\varphi(t)$ : command variable image processing  
 $\delta(t)$ : control variable image processing  
 $e(t)$ : control deviation robot position controller  
 $e_b$ : minimum position error of the control  
 $e_{b,g}$ : response time of the control until minimum position error is reached  
h: height  
 $G_R$ : controlled System  
 $K_{PR}$ : proportional element  
 $K_{IR}$ : integral element  
 $K_{DR}$ : differential element  
s(t): manipulated variable in position control  
t: time  
w: width  
y(t): control variable position controller  
 $y_m(t)$ : overlay position of robot  
w(t): command variable robot position controller

## 2. Tracking approaches for flexible robot guidance

The usage of conveyor belt systems in robot-based industrial automation enables an efficient workflow. The combination of a machining and transport process brings an additional enhancement for mastering complex force-torque controlled assembly tasks in industry. Unpredictable forces and torques often occur during assembly tasks due to modeling inaccuracies and tolerance fluctuations in components and their positioning, which can lead to instabilities and thus to possible damage or quality defects. [2] A pure position-controlled robot is inadequate for successfully executing these processes.

There are various approaches for carrying out machining processes during movement. One strategy necessitates comprehensive analysis and subsequent synchronization of the robotic motion with the movement of the conveyor, considering parameters such as belt speed, component position, and robot dynamics. [3] If mathematical inaccuracies occur due to unpredictable behavior of the conveyor system, such as jerky movements of the conveyor system, this leads to a discrepancy between the planned feed rates of the robot movement and the actual position of the component.

Another approach to conveyor tracking in robotics is the use of camera systems, which is known as visual servoing. In the field of visual servoing, the robot arm can be used as a flexible system to guide the camera along the conveyor.

A second variant is the stationary mounting of the camera system, whereby the robot's working area is covered by the camera's field of view. [4, 5]

Both approaches require object recognition as a first step. Then, the position of the object is determined with the help of the image data. In order to track the moving object and perform assembly tasks, this processing pipeline must be executed continuously. Various approaches are listed in the literature that use 3D camera technology to recognize the position of objects. In most cases, the CAD model is used to achieve a high degree of accuracy. [6, 7] New industrial solutions for flexible robotic

guidance like GuideNOW from inbolt achieve accuracies of up to  $\pm 1$  mm. [8]

Another approach is based on pose estimation without 3D models. In [9] the general framework BundleTrack is used for 6D pose tracking of novel objects with a cycle time of up to 100 ms. Approaches for recognizing the position of previously trained objects from 2D video images are increasingly being pursued. [10, 11] To achieve higher performance, the image information is continuously reduced to keypoints. An evaluation on physical robots, including real-time position control, is not considered in any of the papers mentioned.

## 3. System setup

As stated previously, defined features on the component, workpiece carrier or belt that represent the movement of the conveyor, are analyzed using image processing and integrated into the control loop of the robot path planning as control value. [12] This method requires an image processing system to detect the position of the conveyor, regardless of the actual application, and under various challenges such as feature visibility and fluctuating lighting conditions.

The system set up shown in Fig. 1 therefore consists of a 6-axis NC-robot Comau Racer-7-1.4, a monochrome image processing camera with a sensor width of 1280 pixels and sensor height of 960 pixels, the semiconductor laser distance sensor optoNCDT 1420 with a wavelength of 670 nm and an XY axis portal for moving the component to be tracked and an evaluation pyramid on a workpiece carrier. The laser spot is located on the upper center of the evaluation pyramid with the aim to evaluate the translative tracking performance. The tool coordinate system  $T_{\text{tool}}$  is determined from the robot's world coordinate system  $T$  using coordinate transformation following the described approach in [13]. The XY plane of  $T_{\text{tool}}$  is aligned parallel to the component to be tracked.

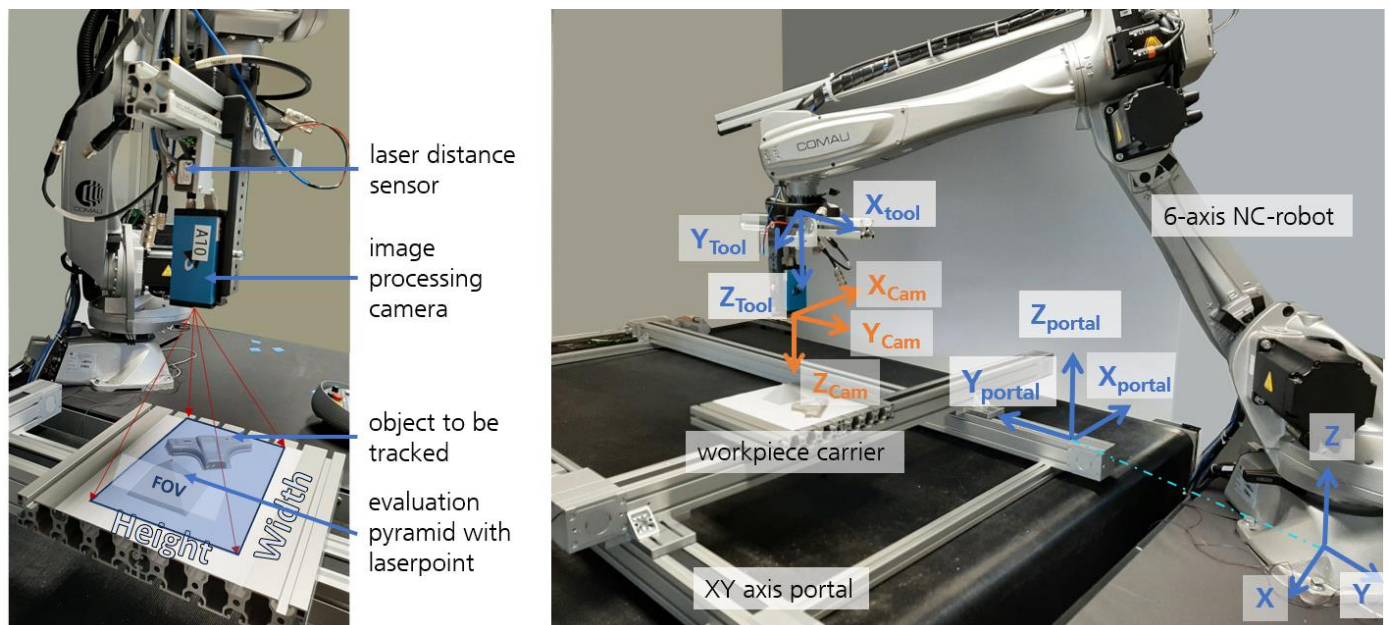


Fig. 1. Schematic overview of the test system for object tracking with illustration of coordinate systems.

The image processing camera is attached to the robot's flange so that the camera axis is aligned perpendicular to the component plane. The origin of the camera coordinate system  $T_{cam}$  is derived from  $T_{tool}$  via a rotation around the Z axis and a translation in X and Y.  $T_{cam}$  is used as a reference for the image processing algorithms. The transformation between T and the coordinate system of the XY axis portal  $T_{portal}$  is used to draw conclusions about the performance of tracking algorithms carried out by the robot that does not communicate with the portal.

#### 4. Development rule-based object tracking algorithms

##### 4.1. Image processing

To ensure a precise tracking, the first and most important step is to correct the barrel lens distortion. This is achieved by calibrating the camera using a chessboard to determine the intrinsic camera parameters and the distortion coefficients. These parameters are then used to map pixels to millimeters.

In the next step ECC (Enhanced Correlation Coefficient) is used to determine the translative and rotational position offset. ECC is an image alignment algorithm that takes two images as input and calculates the transformation between them. The first image is the so-called reference image or initial image (represented as a vector  $\vec{i}_r$ ). The second image is the live image from the camera stream (represented as a warped vector  $\vec{i}_w(\mathbf{p})$ ). In summary, ECC attempts to find a mapping between  $\vec{i}_r$  and  $\vec{i}_w(\mathbf{p})$  under consideration of the following criterion:

$$E_{ECC}(\mathbf{p}) = \left\| \frac{\vec{i}_r}{\|\vec{i}_r\|} - \frac{\vec{i}_w(\mathbf{p})}{\|\vec{i}_w(\mathbf{p})\|} \right\|^2 \quad (1)$$

Based on this approach and its solution, presented in [14], all necessary information can be calculated from the parameter vector  $\mathbf{p}$ .

When considering the coordinate system of the image processing camera and its field of view (height and width) the 6D position offset for the robot  $\delta_{img}(t)$  can be determined. For the use case of simple tracking of a moving object on a conveyor belt, only the translation in X and Y as well as the rotation around the Z axis are of importance. Therefore, neither the rotations around the X and Y axes nor the Z translation are included in the evaluation. This allows the 6D problem to be simplified to one with 3 degrees of freedom, which leads to an optimization of the computing time and an improvement of the energy performance.

As represented in Fig. 2, the camera axis with its relevant coordinates ( $X'_{cam}$ ,  $Y'_{cam}$ ,  $\alpha'_{cam}$ ) is perpendicular to the FOV of the camera or object plane and lies in the center of the target image. The FOV coordinate system  $T_{FOVoffset}$  for determining the offset values between the reference and live image is introduced by shifting the camera coordinate system to the top left corner of the image in the object plane.

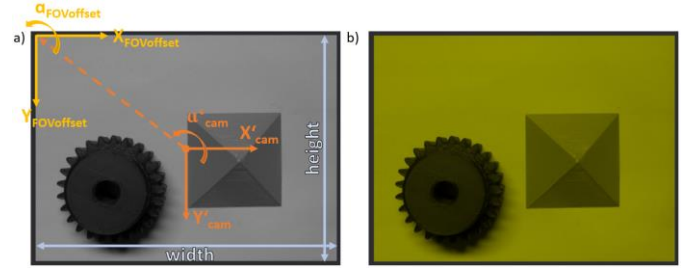


Fig. 2. (a) reference image for gear tracking including coordinate systems of camera and FOV for offset calculation; (b) live image corresponding to target image

If the live image corresponds to the reference image as in Fig. 2, the overlapping area is highlighted yellow. A deviation or offset, as shown in Fig. 3, is highlighted in green. Since  $T_{FOVoffset}$  orientation is the same as  $T_{robot,tool}$ , the three relevant deviations  $\Delta X_{robot}$ ,  $\Delta Y_{robot}$ ,  $\Delta \alpha_{robot}$  can be sent to the robot controller via a TCP connection and processed into relative movements. To ensure a precise movement in this study a calibrating procedure referring to hand-eye calibration is used. [15] This approach helps to determine the transformation matrices from the world coordinates T to the camera  $T_{cam}$ .

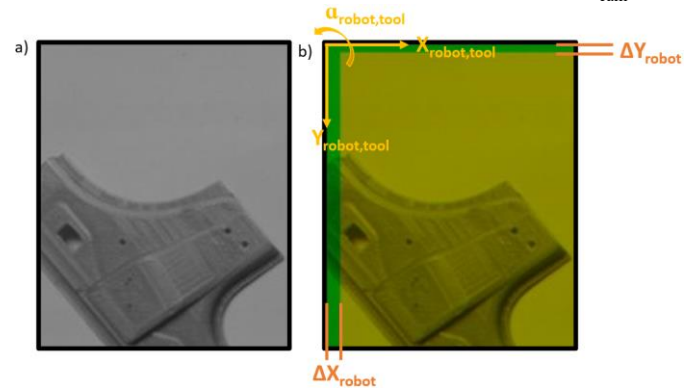


Fig. 3. (a) reference image for tracking of b-pillar (grey); (b) live image in yellow with the position deviations highlighted in green and the position offset in X, Y and C

By adjusting the camera zoom and specifying the width and height of the FOV, the image capture position is flexible. This relation can also be used to enable tracking in Z. However, the reference image must be adjusted with the FOV of the camera depending on the distance.

The developed image processing pipeline is represented as a network in the Xeidana® development environment [16] and is shown in Fig. 4. In addition, interface and necessary preprocessing modules such as image size reduction and lens correction have also been implemented, whereby it should be noted that the accuracy of the position calculation depends directly on the pixel size.

Since no pre-training is required when using ECC, and the reference image is sufficient, it is also referred to as a one-shot demonstration.

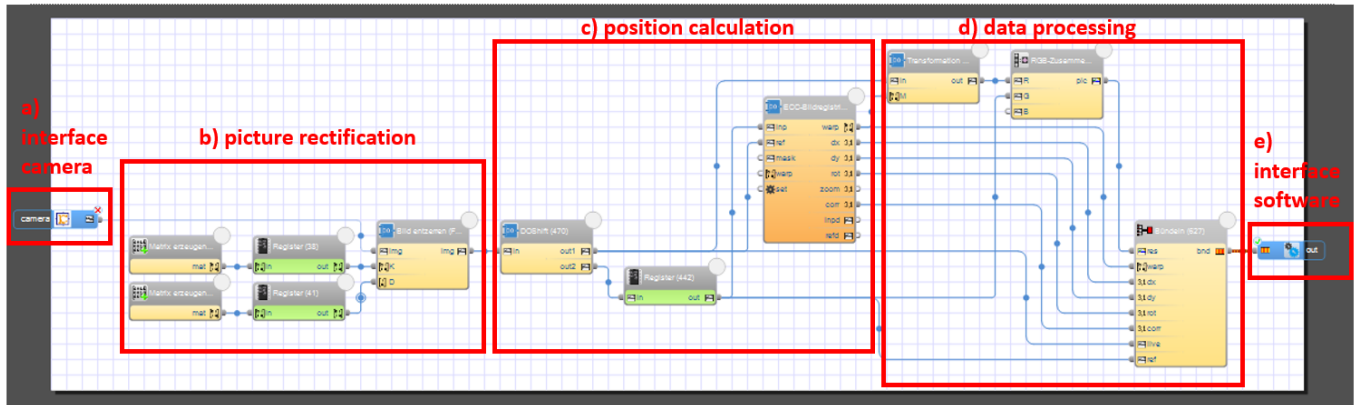


Fig. 4. Developed image processing pipeline consisting of main modules a) interface camera b) picture rectification c) position calculation d) data processing e) interface software, represented as network in the Xeidana® framework, developed by Fraunhofer IWU [16]

#### 4.2. Tracking pipeline and position control

The tracking pipeline, shown in Fig. 5, consists of the two main structures:

- Image processing with calculation of the position deviation running on an eight-core Intel Xeon E-2286 M processor with maximum utilization of 15 % for calculation operations at 15 fps
- The robot position controller on SINUMERIK 840Dsl working in interpolation cycle (IPO) of 4 ms

On the first run, the reference image is captured and saved with its reference position. Since the live image will be taken after position control, the robot does not carry out any movements. With the actual live image, the offset values are calculated as described previously and saved in  $\delta_{img}(t)$ . The absolute position of the robot is then calculated and the new command value for position control  $w(t)$  is set. The position control of the robot is realized by means of synchronous actions which are calculated at 250 Hz and is aimed at calculating the step size for robot movement in IPO-cycle. A PID controller with the following structure is implemented for this purpose.

$$G_R = K_{PR} + \frac{K_{IR}}{s} + K_{DR}s \quad (2)$$

The proportional  $K_{PR}$ , the integral  $K_{IR}$  and the differential component  $K_{DR}$  must be set as control parameters depending on the system behaviour. In reality, the continuous laplace variable  $s$  corresponds to a discrete variable with a time interval of 4 ms. After the PID control is executed, the new actual position of the robot (overlay position) and the remaining offset to  $\delta_{img}(t)$  are calculated using the executed size of the robot. The result is then the new command value for position control in the next loop.

In addition, the control is embedded in a safety circuit that monitors the maximum step and the robot's workspace as well as the correlation between the live and reference image. If the limit values are exceeded, the program is either stopped or the target and axis positions are reset.

#### 5. Experiments and results

In order to design the PID controller, a 2D dimensional movement profile ( $X, Y$ ) was programmed on the axis portal, which accelerates the component to a standard industrial conveyor belt speed of 1 m/min. Based on an identical starting and end position, Figure 6 illustrates the relative movement profile of the portal in  $X$  over time, the associated positions from image processing and the absolute position of the robot.

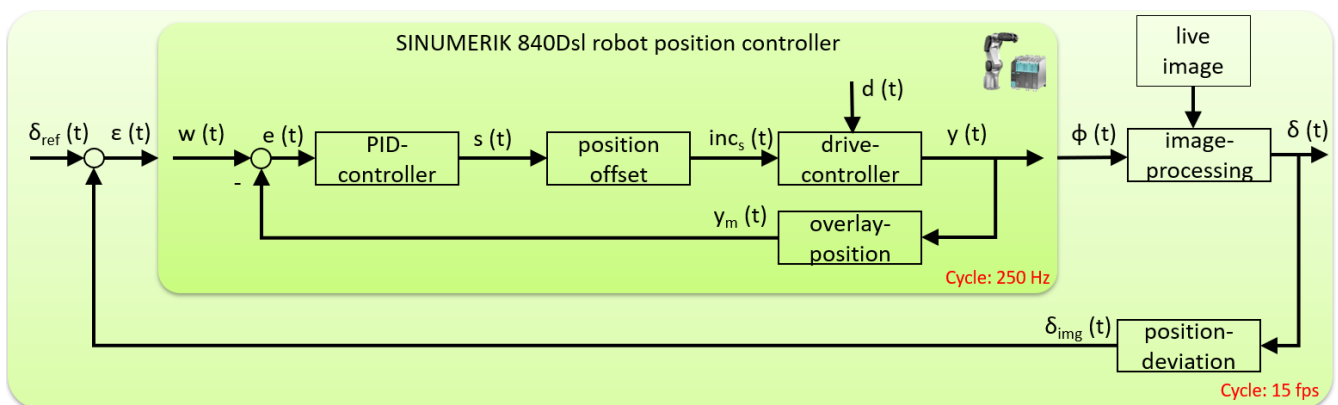


Fig. 5. Processing pipeline and control structure for object tracking consisting of SINUMERIK 840Dsl robot position controller, image processing and position deviation calculation

Table 1. Minimum position error  $e_b$  and velocity error  $e_{b,g}$  for position control of an uniformly moving object at  $1 \frac{m}{min}$  for various controller parameters

Type of controller	$K_{PR}$	$K_{IR}$	$K_{DR}$	$e_b$ [mm]	$e_{b,s}$ [s]
P1	$P_{XY} = 0.007; P_C = 0.007$	-	-	$3.60 < e_b < 4.40$	5.12
P2	$P_{XY} = 0.01; P_C = 0.007$	-	-	$-5.10 < e_b < 7.90$	$\infty$
PID1	$P_{XY} = 0.005; P_C = 0.001$	$I_{XY} = 0.003; I_C = 0.0005$	$D_{XY} = 0.001; D_C = 0.0001$	$0.36 < e_b < 0.80$	13.62
PID2	$P_{XY} = 0.007; P_C = 0.001$	$I_{XY} = 0.003; I_C = 0.0005$	$D_{XY} = 0.001; D_C = 0.0001$	$0.39 < e_b < 0.80$	8.95
PID3	$P_{XY} = 0.009; P_C = 0.001$	$I_{XY} = 0.003; I_C = 0.0005$	$D_{XY} = 0.001; D_C = 0.0001$	$0.12 < e_b < 0.85$	13.41
PID4	$P_{XY} = 0.007; P_C = 0.001$	$I_{XY} = 0.005; I_C = 0.0005$	$D_{XY} = 0.001; D_C = 0.0001$	$0.38 < e_b < 0.60$	11.44

A total of five approximately uniform movement sections, recognisable by the discontinuities (equivalent to abrupt changes in direction of the workpiece carrier), can be distinguished.

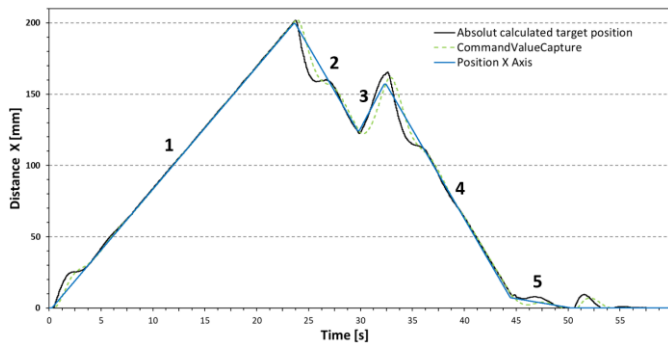


Fig. 6. Performance of the developed tracking pipeline consisting of *CommandValueCapture* from image processing and absolute target position of the robot compared to absolute X target position of the X-portal at  $1 \frac{m}{min}$

By looking at the first time section in X direction ( $t < 15$  s) the processing pipeline including different PID controllers are analysed. Table 1 gives an overview of them as well as of the position control error  $e_b$  and the response time  $e_{b,s}$  until the position error is minimized.

The first two P controllers (P1 and P2) underline clearly that a persistent disturbance variable cannot be precisely controlled in a stationary manner for a system with an equalization. However, the residual control error cannot become completely zero with a finite controller gain. A maximum gain  $K_{PR} = 0.007$  with  $e_b$  of approx. 4 mm was empirically determined. In addition, an oscillation can be recognized if the gain is too high (e.g. P2). This contradiction can only be resolved if the controller can generate a sustained manipulated variable even though the control error has become zero. For this purpose, controllers with an I component ( $K_{IR}$ ) are able to do this due to their "memory effect". However, these systems are more inert, which is why a D component ( $K_{DR}$ ) is considered.

The PID controllers tested for a movement in X of the axis portal are presented in Fig. 7. The following graphs and statements can be transferred analogously to translative movements in Y.

The offset curves of the position control over time show that the control speed  $e_{b,s}$  becomes smaller when a larger  $K_{PR}$  is selected (comparison between PID1 and PID2), however, no significant improvement for  $e_b$  is achieved.

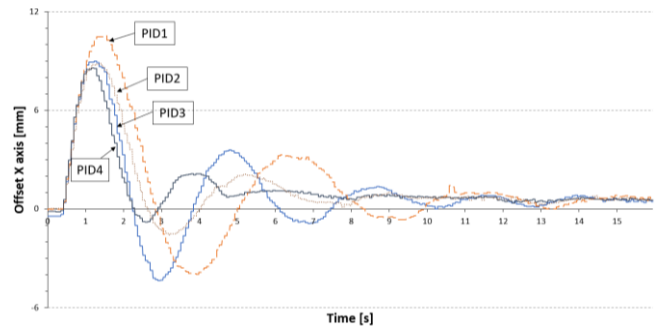


Fig. 7. Comparison of PID controllers 1-4 for position control in X direction for an object velocity of  $1 \frac{m}{min}$

With higher gain factors, the system tends still to oscillate and thus becomes more unstable (PID3). This is also accompanied by a deterioration in  $e_{b,s}$ . Increasing the I component improves the accuracy of the control, while at the same time increasing the delay (PID4). Although the use of the D component counteracts this inertia, higher  $K_{DR}$  values lead to an even higher oscillation of the system, especially at the discontinuities. It can be concluded that PID2 performs best, noticing it reacts quickly to changes in speed with the same accuracy as the other controllers. The mostly sufficient accuracy of  $\pm 1$  mm compared to existing tracking solution with 3D cameras is achieved after approx. 6 s.

The dependencies of the absolute calculated target position of the robot, the command values of image processing and position control as well as the performed robot step in IPO-Cycle for the first 0,3 s from beginning of tracking operation, are illustrated in Fig. 8.

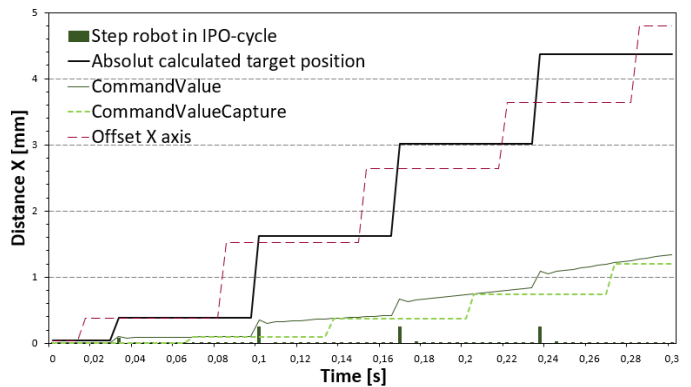


Fig. 8. Robot position control in X direction showing the actual position offset of the robot to the object in the first 0,3 s from receipt of a disturbance (object is accelerated to a speed of 1 m/min)

It can be concluded that the processing time of the ECC is the limiting factor in the processing pipeline. The position controller always reaches the target position at the achieved image processing speed. By reducing the data to be analyzed to 10 % of the available data, a processing rate of 15 fps could be achieved with 10,5 W of power consumption on eight-core Intel Xeon E-2286 M processor.

## 6. Conclusion and further development

The overarching goal of this study will be a streamlined process that enables on-the-fly assembly through the intelligent application of combined vision and force-torque data, negating the need for additional sensory equipment.

The basis for this was presented in this paper. Based on the ECC approach, an NC robot can track the movements of any objects for 3 degrees of freedom according to one shot demonstration with an accuracy of <1 mm. Particularly noteworthy is the low energy consumption of 10.5 W for the calculation algorithms and the low utilization of a commercially available PC.

Further improvements could be achieved by using more powerful hardware or by further data reduction methods (smaller image section, lower resolution, etc.). However, these approaches are either at the expense of energy consumption or the accuracy of the control. It should also be noted that image processing always provides positions from the past to the robot controller. To minimize the position error of the control, approaches for predicting positions in the future for at least uniform movements, such as the additional use of a Kalman filter, will be tested. Furthermore, the dependency of the position control on the image processing speed in combination with the energy consumption will be further investigated to improve tracking.

These improvements are intended to achieve application scenarios where robots are equipped with cameras and force-torque sensors and where contact between the robot and the component occurs due to the application. Consequently, the tracking approach presented in the paper must be extended to include movements in the Z direction, using a single 2D camera with the aim that the robot approaches the moving object at a defined point until contact occurs. At this point, the system switches from visual servoing to force torque control to perform the assembly task. The innovative contribution of next studies is to expand stationary implementation of the assembly tasks, like described in [17], to an on-the-fly assembly.

## Acknowledgements

This work has been accomplished with financial support by Federal Ministry of Economic Affairs and Climate Action of Germany and the under contract Grant 01MJ22001A and 01MJ22001B “GreenBotAI - Frugal and adaptive AI for flexible industrial Robotics”.

## References

- [1] GreenBotAI- Frugal and adaptive AI for flexible industrial robotics, Fraunhofer IWU, Chemnitz, [2023], online ressource, URL. <https://greenbot-ai.eu/>, last viewed at 29.04.2024
- [2] Wolfgang Weber, Heiko Koch, *Industrieroboter Methoden der Steuerung und Regelung*; 2022 Carl Hanser Verlag Munchen, pp. 207-209
- [3] T. H. Park and B. H. Lee, "An approach to robot motion analysis and planning for conveyor tracking," *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, Sacramento, CA, USA, 1991, pp. 70-75 vol.1, doi: 10.1109/ROBOT.1991.131555.
- [4] G. Flandin, F. Chaumette and E. Marchand, "Eye-in-hand/eye-to-hand cooperation for visual servoing," *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, San Francisco, CA, USA, 2000, pp. 2741-2746 vol.3, doi: 10.1109/ROBOT.2000.846442.
- [5] F. Reyes and R. Kelly. Experimental evaluation of fixed-camera direct visual controllers on a direct-drive robot. In *IEEE Int. Conf on Robotics and Automation*, Leuven, Belgium, May 1998.
- [6] H. Wang et al., "Normalized object coordinate space for category-level 6D object pose and size estimation", *CVPR*, 2019.
- [7] Zhang X, Lv W, Zeng L. A 6DoF Pose Estimation Dataset and Network for Multiple Parametric Shapes in Stacked Scenarios. *Machines*. 2021; 9(12):321. <https://doi.org/10.3390/machines9120321>
- [8] GuideNOW brochure downloadable at <https://www.inbolt.com/applications>
- [9] B. Wen and K. Bekris, "BundleTrack: 6D Pose Tracking for Novel Objects without Instance or Category-Level 3D Models," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 2021, pp. 8067-8074, doi: 10.1109/IROS51168.2021.9635991.
- [10] Y. Lin, J. Tremblay, S. Tyree, P. A. Vela and S. Birchfield, "Keypoint-Based Category-Level Object Pose Tracking from an RGB Sequence with Uncertainty Estimation," *2022 International Conference on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, 2022, pp. 1258-1264, doi: 10.1109/ICRA46639.2022.9811720.
- [11] C. Wang et al., "6-PACK: Category-level 6D Pose Tracker with Anchor-Based Keypoints," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 2020, pp. 10059-10066, doi: 10.1109/ICRA40945.2020.9196679.
- [12] Bruno Siciliano, Oussama Khatib (Eds.), *Handbook of Robotics*, Springer-Verlag Berlin Heidelberg 2008 ; pp.563-582
- [13] Zhe Han, Huanyu Tian, Tom Vercauteren, Da Liu, Changsheng Li, Xingguang Duan, Collaborative human-robot surgery for Mandibular Angle Split Osteotomy: Optical tracking based approach, *Biomedical Signal Processing and Control*, Volume 93, 2024, 106173, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2024.106173>.
- [14] G. D. Evangelidis and E. Z. Psarakis, "Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858-1865, Oct. 2008, doi: 10.1109/TPAMI.2008.113.
- [15] Roger Y. Tsai, Reimar K. Lenz, A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration, *IEEE transactions on robotics and automation*, vol. 5, no. 3, june 1989.
- [16] XEIDANA®: Neue Software für die automatisierte Qualitätssicherung, Fraunhofer IWU, Chemnitz, [2024], online ressource, URL. <https://xeidana.de>, last viewed at 15.04.2024
- [17] Yunqi Gu, Ruth Maria Otto, Martin Naumann, Leutrim Gjakova, Rico Löser, AI-driven force torque control strategies for further automate flexible high-precision, contact-intensive assemblies, to be published at 57th CIRP Conference on Manufacturing Systems 2024 (CMS 2024), Povia de Varzim, Portugal, 2024