Proceedings of the Canadian Society for Mechanical Engineering International Congress
32nd Annual Conference of the Computational Fluid Dynamics Society of Canada
Canadian Society of Rheology Symposium
CSME-CFDSC-CSR 2025
May 25–28, 2025, Montréal, Québec, Canada

AI-Enabled Machine Vision Model for Manual Process Monitoring and Cycle Time Measurement

Daniela Sanchez[*, 1], Sajid Ullah Butt[1], Ahmed Qureshi[1]
[1]Mechanical Engineering Department, University of Alberta, Edmonton, Canada.
*desanche@ualberta.ca

Abstract—In high-variety, high-volume production, optimizing manual assembly is essential to sustain productivity, ensure flexibility, and minimize inefficiency. Unlike automated systems, manual assembly relies on human operators performing complex, repetitive tasks, which are prone to variations in execution time, skill level, and consistency. Traditional cycle time tracking methods, often manual and error-prone, lack precision and granularity. This research explores AI-powered machine vision to automate action recognition and cycle time measurement, offering a scalable, data-driven solution for real-time process optimization.

By leveraging advanced video analytics, the system identifies and classifies operator actions, recognizes key process steps, measures cycle times accurately, and detects inefficiencies. These insights enable precise time studies, helping manufacturers pinpoint bottlenecks, streamline workflows, and enhance efficiency. Unlike traditional sensor-based methods, AI-powered machine vision provides non-intrusive, adaptable tracking, making it ideal for dynamic production environments.

Integrating machine vision aligns with digital manufacturing principles, using data-driven approaches for continuous improvement and quality control. The system delivers real-time insights and predictive analytics, enabling proactive adjustments to production planning and resource allocation. This approach strengthens flexible manufacturing, allowing rapid adaptation to product variability and shifting demands. Ultimately, the solution equips manufacturers to manage volume and diversity fluctuations without compromising quality or throughput, enhancing productivity and competitiveness.

Keywords- AI in manufacturing; Industry 4.0; Smart manufacturing, Digital manufacturing, flexible manufacturing

I. Introduction

In today's competitive landscape, industries must continually adapt to technological advancements in manufacturing to remain relevant. Industry 4.0, driven by the integration of artificial intelligence (AI), the Internet of Things (IoT), big data, and cloud computing, has marked the beginning of a new era in manufacturing. These technologies have revolutionized traditional manufacturing methods, significantly improving efficiency, productivity, and adaptability [1].

Manufacturing processes often involve repetitive tasks performed by both machines and human operators. While these tasks are flexible and adaptable to varying product specifications, they are prone to inconsistencies and inefficiencies caused by machine variation, accuracy limitations, human error, and differences in skill levels. Although mechanized processes can be effectively monitored using sensor fusion, manual processes pose challenges in gathering comprehensive process data through traditional methods.

This paper aims to address these limitations by leveraging AI-driven video analytics to analyze real-time video data from manual processes in the assembly lines. The proposed system seeks to identify key process steps, monitor cycle times, and enhance quality control. By automating work studies and cycle time analysis, the objective is to provide real-time insights that drive continuous improvement and operational optimization in manual assembly lines.

II. Literature Review

Industry 4.0 has led to the development of more intelligent and adaptive systems, allowing manufacturing environments to become dynamic and interconnected, providing real-time data acquisition and analysis [2]. As a result, companies can make informed decisions, optimize operations, and quickly adapt to changes in demand and production conditions.

At the heart of this transformation are emerging technologies like artificial intelligence (AI), the Internet of Things (IoT), big data, and cloud computing. Together, these technologies enable a level of connectivity and automation, bringing innovative manufacturing paradigms such as Smart Manufacturing (SM) and Intelligent Manufacturing (IM) [3].

Smart manufacturing is defined as a collection of various technologies that can promote a strategic innovation of the existing manufacturing industry through the convergence of humans, technology, and information [2]. SM focuses on creating systems that are not only automated but also capable of self-optimization, predictive maintenance, and autonomous decision-making to increase productivity and reduce error and waste.

Moreover, Intelligent Manufacturing builds upon SM by focusing on embedding AI and machine learning into manufacturing systems. IM emulates human-like decision-making, enabling manufacturing operations to adapt to complex and unpredictable scenarios [1]. For example, an IM-enabled system uses real-time data to dynamically make use of resources that not only improve productivity and product quality but also drive innovation and long-term sustainability.

In the industrial context, AI is a key concept that enables smart and intelligent manufacturing capabilities. Industrial AI focuses on five key dimensions: **Infrastructures**, which prioritize real-time processing, reliability, security, and interconnectivity; **Data**, characterized by its large volume, high velocity, and variety from diverse sources; **Algorithms**, integrating physical, digital, and heuristic knowledge to manage complex models; **Decision-making**, where low tolerance for error and efficiency in handling uncertainty is critical for optimizing large-scale industrial problems; and **Objectives**, which are centered on concrete value creation through improvements like scrap reduction, quality enhancement, operator performance, and faster ramp-up times [4].

AI techniques like machine learning and deep learning enhance decision-making, and production management across various stages of the industrial lifecycle, including design, manufacturing, logistics, and Maintenance.

Machine vision is the automatic acquisition and analysis of images to obtain desired data to control a process or activity [5]. For manual processes, machine vision cameras and specialized sensors capture detailed images of production activities, while machine learning techniques enable the identification and classification of components based on distinct visual features. When integrated with machine learning, machine vision systems can accurately classify, detect, and track desired objects. Key applications in manufacturing include:

A. Process monitoring:

Machine vision systems can measure operational parameters in real-time. For instance, Lou et al. [6] proposed a vision system that integrates a contactless monitoring framework utilizing machine vision to enable real-time observation and counting manual operations in production and assembly lines. The system employed YOLOv4 to identify actions and count each instance of a task being completed. By tracking repetitions, the system provided insights into the number of tasks an operator performed within a given timeframe, enabling comparisons of performance across shifts, workers, or variations in tasks.

B. Defect Detection:

Early defect detection ensures that faulty products are identified before further processing. One of the studies designed a machine vision system for identifying geometric anomalies in components, reducing the rework costs significantly [7]. Moreover, YOLOv5 is used in the food industry to detect kiwifruit defects, showing the versatility of machine vision in automatic defect detection in many domains [8].

C. Quality Control:

In an assembly line, machine vision systems enable seamless inspection without interrupting the continuous production flow or requiring periodic halts [9]. These systems are not only immune to human errors but also excel in speed and precision and process each image in mere milliseconds while achieving high levels of accuracy [10].

D. Human Action Recognition:

Recent advancements enable machine vision to monitor the activities of one or more individuals through the analysis of video sequences, offering a non-intrusive alternative to wearable sensors for monitoring and evaluating activities on the production floor [11][12]. Technologies such as machine vision and human action recognition are central to modern advancements in manufacturing, offering non-intrusive, real-time monitoring and quality control capabilities. In the context of an assembly line, particularly for manual processes, video-based data provides critical insights into the tasks being performed. For instance, it can track an operator assembling a product through a sequence of images, identifying specific tools and components involved in the process.

Prior research has successfully utilized YOLO to find defects, perform quality checks, and recognize actions. Though these are beneficial applications in manual assembly lines where operators work in parallel, there is a lack of research addressing the complexities of tracking and analyzing concurrent actions. This impairs the potential for cycle time calculation and process optimization opportunities.

By using an AI-driven system to measure time and monitor manual assembly operations, this research aims to bridge the gap between real-time action recognition and lean manufacturing objectives ultimately improving productivity and process flow in manual assembly lines.

III. PROBLEM STATEMENT

Efficiency, productivity, and adaptability are key issues in manufacturing industries. Although automation has aligned many processes, manual assembly lines play a vital role in high-mix, high-volume production environments because of their flexibility in handling diverse product configurations.

Product variations in high-mix production line significantly influence manual process times resulting in inefficiencies, inconsistent quality, and missed opportunities to enhance productivity. Without a comprehensive understanding of operator actions and their impact on production cycles, manufacturers struggle to optimize workflows and improve performance.

This research seeks to address these challenges by developing an AI-driven video analytics system that identifies and classifies operator actions in real-time. The system will analyze video data from manual assembly lines, calculate cycle times for various product configurations, and provide actionable insights to optimize operator motions, streamline workflows, and enhance overall efficiency

IV. METHODOLOGY

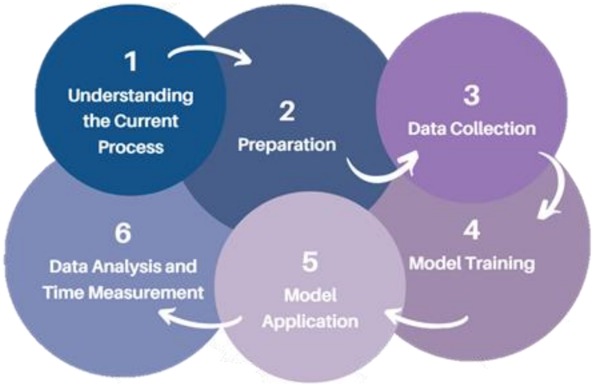

Figure 1. Proposed methodology for process monitoring

The steps involved in the proposed methodology are shown in Figure 1 and are detailed below:

1) *Understanding the Current Process.*

The first phase focuses on understanding the existing manual assembly line workflow. This involves visiting the shop floor to observe the process and gain insight into how tasks are executed. As part of this phase, the assembly process is broken down into distinct actions, identifying the operations sequence and documenting how operators distribute tasks.

2) *Preparation*

Since the assembly line involves parallel operations, this phase focuses on identifying the key features and components of each action. By clearly defining key actions, each action can be further analyzed and broken down into its specific components, enabling a detailed understanding of the process. In this study, each action is taken as a group of three components: the part being assembled, the tool used for assembly, and the operator performing the action, as shown in Figure 2.

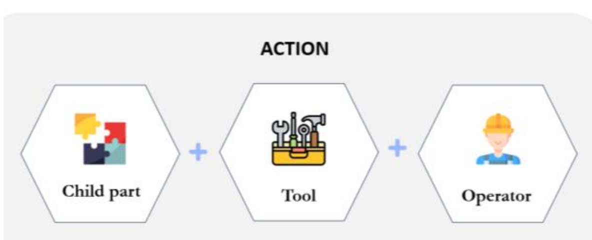

Figure 2. Action as a combination of three Components

3) *Data Collection*

To train the AI model, this phase gathers the process data by capturing the videos and images of operators performing each action along with specific tools and sub-parts. Based on the data, it builds a dataset of the components involved in the process. For the robustness of the collected data, the angles, illumination conditions, and other operating conditions are varied during data collection.

TABLE I. TOTAL DEFINED LABELS PER CLASS

Labels	***Labeled dataset***
Part_1	850
Part_2	446
Tool_1	579
Tool_2	542
Tool_3	146
Total	2563

4) *Model Training*

Using the collected data, images are annotated with labels corresponding to the actions and components identified in the workflow analysis (Table I). A YOLOv8n model, a convolutional neural network (CNN)-based deep learning model specialized in real-time object detection, is then trained to recognize individual components under varying conditions. Following this, an action logic is implemented to validate whether there is an overlap between the detected labels. This enables the classification of actions within the assembly process, as shown in Table II.

TABLE II. DEFINED LOGIC FOR ACTIONS

Action	***Labels***
Action 1	Tool_3
Action 2	Visual Validation (not in model)
Action 3	Part_1
Action 4	Part_2
Action 5	Part_2 + Tool_1
Action 6	Part_1 + Tool_2

5) *Model Application*

Once the model is trained to recognize different actions, it is applied to other captured process videos of the assembly line to detect and classify actions in real time. The code is designed to track and count detected actions, enabling the calculation of the time taken for each action by identifying the first and last frames in which the action is recognized.

6) *Data Analysis and Time Measurement*

The model was applied to 11 videos from products with different configurations, the time measured is shown in Table III.

TABLE III. MODEL APLICATION RESULTS

Video ID	***Total Frames***	***Total cycle Time (sec)***	***Actions Counts***
1	3752	116.57	4
2	3929	148.07	6
3	3965	111.29	4
4	3788	109.78	6

Video ID	*Total Frames*	*Total cycle Time (sec)*	*Actions Counts*
5	4177	161.47	6
6	4744	160.63	6
7	1888	34.20	4
8	3965	155.98	6
9	5098	256.85	6
10	3575	145.07	6
11	3304	50.68	6

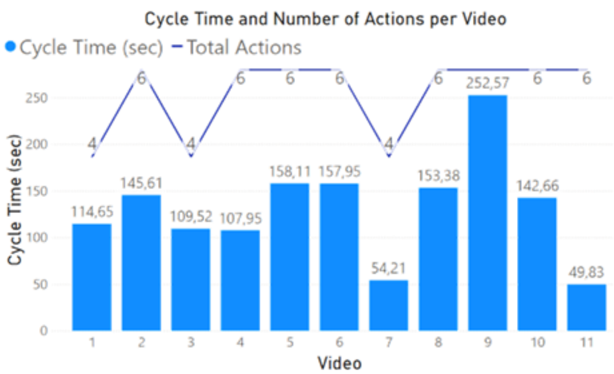

Figure 3. Total Cycle Time and Actions counts by Video

Figure 3 summarizes the initial results, highlighting variations in total cycle times and action counts across the 11 analyzed videos. These differences are likely due to variations in product type, size, and configuration in each video. For instance, Videos 1, 4, and 7 exhibit shorter cycle times and fewer recorded actions, as certain steps such as Actions 4 and 5 were not performed.

Additionally, the videos were recorded on different days, from varying angles, and under different conditions, which may also impact on the model's ability to accurately classify actions.

A deeper validation was conducted by printing the frames where each action was identified (as shown in figure 4) to assess the continuity of action recognition. Video 3 was used as a reference for this exercise, aiming to analyze the model's detection of actions, evaluate recognition consistency, and identify any gaps between the initial and final detection of each action. This process helps refine the model's accuracy in tracking actions and ensures a more precise measurement of task durations.

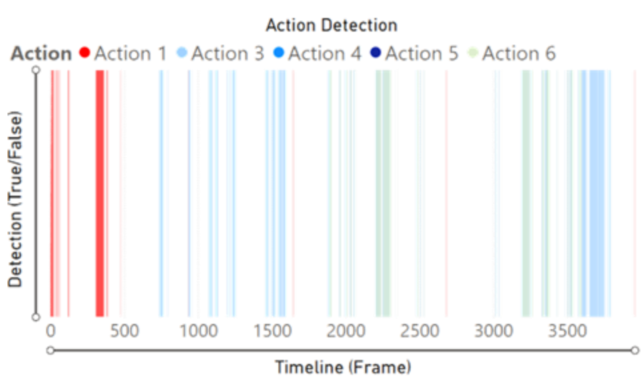

Figure 4. Action recognition by frames - Video #3

While all actions were analyzed, particular focus was placed on Action 1 due to its previous misclassification issues involving other objects. By isolating the action logic for this specific task, it became evident that the recognition of tool_3 was inconsistent, with significant gaps between detected frames.

To enhance the reliability of action classification, Action 1 was redefined and linked to an additional object. This adjustment ensured that the model verifies the presence of both objects, assesses their overlap, and accurately classifies the action, ultimately improving detection accuracy and reducing misclassification errors.

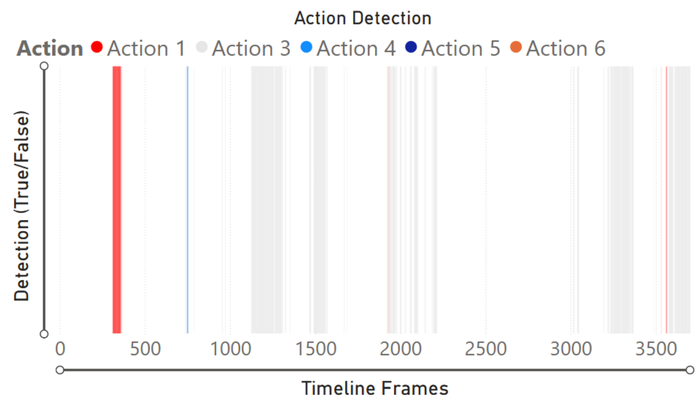

Figure 5. Action recognition by frames - Video #3 refinement

As highlighted in orange in Figure 5, this new approach significantly improved the model's ability to identify and classify Action 1 compared to the previous method. However, there is still room for improvement, as misclassifications of this action continue to appear in certain frames toward the end of the video. Table IV provides a detailed breakdown of the specific frames where the action was detected. A manual validation was conducted, identifying a total of 58 frames where Action 1 was visibly present. This confirms that the second approach yielded more accurate and consistent results in detecting this action. However, further refinements are still necessary to enhance detection reliability and minimize errors across the entire sequence.

TABLE IV. ACTION 1 TIME RESULTS

Test	1st frame	Last frame	Total Frames count	Frames with Bounding boxes	Real Count	% Error
1	3	3959	3957	92	58	58.6%
2	313	3567	3255	50		13.8%

V. RESULTS AND DISCUSSION

The initial video data results highlight that time variability, driven by differences in product configurations, underscores the importance of incorporating product configuration data into future analyses. By correlating time and action patterns with specific product types, the system can enhance its predictive capabilities and provide more targeted insights for process optimization.

According to the results, Action 1 and Action 3 are potentially the most time-intensive steps in the process, as they exhibit the highest average durations. However, Manual validation revealed inconsistencies in the model's ability to accurately recognize Tool_3, as other objects are frequently misidentified as Tool_3. Since the action logic identifies Action 1 based on the detection of tool_3, this misclassification inflates the recorded time for Action 1 beyond its actual duration. These issues likely stem from inaccuracies in the labeling during the data preparation phase.

This revelation suggests that the action classification logic should be refined by identifying the overlap of two objects rather than relying on a single object, as is currently the case for Action 1. This adjustment might also apply to Action 3 and 4. Leveraging the overlap of two distinct labels could enhance the model's ability to identify and classify actions accurately.

The next steps will focus on optimizing camera positioning and expanding the identification of components associated with each action to determine whether overlapping classes can enhance classification accuracy and reduce inconsistencies. Additionally, the system should validate that action detections occur continuously, ensuring that brief misclassifications spanning only one or two frames are classified as outliers and be excluded from time measurements. This refinement will not only eliminate minor false detections but also help determine whether an action occurs more than once within the same assembly process, providing deeper insights into process variations.

VI. CONCLUSIONS

This study highlights the potential of AI-driven video analytics in optimizing cycle time measurement and process efficiency in a manual assembly line environment. The initial results emphasize the impact of product configuration variability on time measurements, underscoring the need to incorporate product-specific data into future analyses. By linking time and action patterns to different product types, the system will be able to refine its predictive capabilities and generate more targeted insights for process optimization.

Key findings indicate that Action 1 and Action 3 are among the most time-intensive steps in the assembly process. However, inconsistencies in the model's ability to recognize Tool_3 have led to inflated time measurements due to misclassification errors. This issue stems from limitations in the labeling process, reinforcing the need for a more refined action classification logic. Moving forward, incorporating overlapping object detection rather than relying on single-object recognition is expected to enhance classification accuracy and reduce inconsistencies, particularly for Actions 1, 3, and 4.

To improve the system's reliability, future efforts will focus on optimizing camera positioning and refining action detection criteria to ensure continuous and accurate identification. By addressing misclassification errors and eliminating brief false detections, the model can provide more consistent and precise cycle time measurements. Additionally, this enhancement will help determine whether an action occurs more than once in the same assembly process, offering deeper insights into process variability.

Ultimately, this technology aims to provide an automated approach to measuring cycle time, minimizing human error and improving data accuracy in a complex manual assembly environment. Given the inherent variability of manual processes, achieving a system that reduces inconsistencies and adapts to diverse production scenarios is critical. By continuously refining model accuracy and validation techniques, this AI-driven approach can serve as a powerful tool for driving process efficiency, process monitoring, and supporting continuous improvement initiatives in manufacturing.

REFERENCES

[1] Ding, H., Gao, R. X., Isaksson, A. J., Landers, R. G., Parisini, T., & Yuan, Y. (2020). State of AI-based monitoring in smart manufacturing and introduction to focused section. IEEE/ASME Transactions on Mechatronics, 25(5), 2143–2154. https://doi.org/10.1109/tmech.2020.3022983

[2] Wang, B., Tao, F., Fang, X., Liu, C., Liu, Y., & Freiheit, T. (2021). Smart manufacturing and intelligent manufacturing: A comparative review. Engineering (Beijing, China), 738-757. doi:10.1016/j.eng.2020.07.017

[3] Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., . . . Noh, S. D. (2016). Smart manufacturing: Past research, present findings, and future directions. International Journal of Precision Engineering and Manufacturing-Green Technology, 3(1), 111-128. doi:10.1007/s40684-016-0015-5

[4] Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial artificial intelligence in industry 4.0 - systematic review, challenges and outlook. *IEEE access: practical innovations, open solutions*, 220121-220139. doi:10.1109/ACCESS.2020.3042874

[5] West, P. (2021). Fundamentals of Machine Vision.

[6] Lou, P., Li, J., Zeng, Y., Chen, B., & Zhang, X. (2022). Real-time monitoring for manual operations with machine vision in smart manufacturing. Journal of manufacturing systems, 65, 709-719. doi:10.1016/j.jmsy.2022.10.015

[7] Frustaci, F., Perri, S., Cocorullo, G., & Corsonello, P. (2020). An embedded machine vision system for an in-line quality check of assembly processes. Procedia manufacturing, 211-218. doi:10.1016/j.promfg.2020.02.072

[8] o, J., Qi, J., Zhang, J., Shao, H., Yang, J., & Li, X. (2021). A real-time detection algorithm for kiwifruit defects based on YOLOv5. *Electronics, 10*(14), 1711. doi:10.3390/electronics10141711

[9] F. Frustaci, F. Spagnolo, S. Perri, G. Cocorullo, and P. Corsonello, "Robust and High-Performance Machine Vision System for automatic quality inspection in assembly processes," Sensors, vol. 22, no. 8, p. 2839, Apr. 2022, doi: 10.3390/s22082839.

[10] T. Benbarrad, M. Salhaoui, S. B. Kenitar, and M. Arioua, "Intelligent Machine Vision model for defective product inspection based on machine learning," Journal of Sensor and Actuator Networks, vol. 10, no. 1, p. 7, Jan. 2021, doi: 10.3390/jsan10010007.

[11] A. Mastakouris, G. Andriosopoulou, D. Masouros, P. Benardos, G. -c. Vosniakos, and D. Soudris, "Human worker activity recognition in a production floor environment through deep learning," Journal of Manufacturing Systems, vol. 71, pp. 115–130, Sep. 2023, doi: 10.1016/j.jmsy.2023.08.020.

[12] Sharma, V., Gupta, M., Kumar, A., & Mishra, D. (2021). Video processing using deep learning techniques: A systematic literature review. IEEE access; practical innovations, open solutions, 9, 139489-139507. doi:10.1109/access.2021.3118541