COMPUTER VISION AS A SENSOR: AUTOMATED KPIV MEASUREMENT IN INDUSTRIAL SYSTEMS

Lily Chen
Weng Chin Tang
Kenneth F. Raymundo

Factory Analytics and Operations Excellence – Global Operations
Western Digital Corp., 109 Technology Ave., SEPZ, Laguna Technopark, Biñan, Laguna, Philippines 4024
lily.chen@wdc.com, weng.chin.tang@wdc.com, kenneth.raymundo@wdc.com

ABSTRACT

Many legacy industrial machines remain isolated from digital monitoring systems due to the high cost, complexity, or impracticality of retrofitting them with modern sensors or IoT devices. This creates a critical barrier for manufacturers aiming to adopt Industry 4.0 practices, where real-time data and process visibility are essential. To address this challenge, this study proposes a computer vision-based video analytics system that functions as a virtual sensor to automate the measurement of Key Process Input Variables (KPIVs) from legacy equipment. The system utilizes real-time video feeds captured by strategically positioned cameras and applies deep learning models for object detection, optical character recognition (OCR), and visual state recognition to extract meaningful process parameters from analog interfaces, such as dial gauges, indicator lights, and display panels. These parameters are digitized and streamed to a centralized dashboard for real-time monitoring, historical logging, and performance analytics. Deployment across multiple industrial use cases demonstrated high accuracy, responsiveness, and minimal integration overhead. The results validate the system's effectiveness in providing non-invasive, scalable, and cost-efficient digital augmentation of legacy machines. This approach offers a practical pathway to enhance operational transparency, support predictive maintenance, and accelerate digital transformation without altering existing hardware infrastructure.

1. 0 INTRODUCTION

In the age of smart manufacturing and Industry 4.0, the ability to continuously monitor and analyze Key Process Input Variables (KPIVs) is essential for maintaining optimal production quality and operational efficiency. KPIVs—such as temperature, pressure, cycle times, and component status—offer valuable insights into machine health and process consistency. However, many manufacturing lines still operate using legacy machines that lack modern connectivity and sensing capabilities. In our current manufacturing process, KPIV data is typically manually recorded once per shift by human operators. This method introduces several critical challenges such as: It is time-consuming and labor-intensive, reducing operator productivity. It is prone to human error, potentially compromising the accuracy and consistency of the data. It leads to limited data availability, with delayed and infrequent updates that hinder real-time decision-making and monitoring.

These limitations pose a significant barrier to achieving data-driven manufacturing excellence, especially when upgrading legacy systems with IoT hardware is costly or operationally disruptive. To address this, the present work proposes a computer vision-based system that uses deep learning and video analytics to act as a virtual sensor for non-invasive KPIV measurement. By leveraging existing camera infrastructure or deploying low-cost industrial cameras, this approach digitizes visual cues from machine panels, gauges, and indicators—enabling scalable data collection without retrofitting or altering legacy hardware.

1.1 Video Analytics

Video analytics involves analyzing live or recorded video streams to extract meaningful information. In industrial settings, strategically placed cameras can continuously observe control panels, analog displays, and visual signals. These visual data streams are processed to monitor machine states, detect anomalies, or track process changes. Unlike conventional sensors, video analytics is non-intrusive and adaptable, providing a versatile tool for bridging the digital divide in legacy environments.

1.1.1 Deep Learning

Deep learning, particularly through Convolutional Neural Networks (CNNs), enhances video analytics by enabling automated object detection, optical character recognition (OCR), and state classification. These techniques allow the system to recognize dial positions, read numeric displays, and classify indicator light statuses. Trained on representative industrial datasets, these models ensure reliable performance across diverse equipment types, lighting conditions, and viewing angles.

1.1.2 Computer Vision as A Sensor

By interpreting visual data with deep learning, computer vision can function as a virtual sensor—digitizing KPIVs from legacy machinery without any physical intervention. This enables real-time monitoring of operational parameters such as counts, durations, and visual status indicators. The extracted data is then streamed to a centralized dashboard for visualization, alerting, and analytics—empowering manufacturers to make informed decisions, optimize maintenance schedules, and improve process transparency.
Although computer vision has been widely applied in defect detection, robotic automation, and quality control, its application as a sensing mechanism for KPIV acquisition in legacy environments is still emerging. This study contributes to this area by presenting a deployable and cost-effective solution that digitizes manual processes, increases data granularity, and drives smarter manufacturing with minimal infrastructure changes.

2. 0 REVIEW OF RELATED WORK

In recent years, the convergence of computer vision and industrial automation has sparked interest in non-contact sensing techniques for legacy manufacturing systems. Traditional data acquisition in such environments typically relies on Supervisory Control and Data Acquisition (SCADA) systems and physical sensors. However, these solutions often require significant retrofitting, wiring, and integration costs, making them unsuitable for aging equipment with limited digital interfaces [1].
Previous works have demonstrated the efficacy of computer vision for specific industrial tasks. For example, defect detection and visual inspection using convolutional neural networks (CNNs) have been widely explored to improve quality control in production lines [2]. Robotic vision systems for automated pick-and-place operations have also been implemented using deep learning models such as YOLO and SSD for real-time object localization [3]. However, most of these applications focus on automation tasks involving discrete events or quality checks, rather than continuous process monitoring.
Optical Character Recognition (OCR) has been applied to digitalize analog meters and handwritten records in industrial settings. In [4], a deep learning-based OCR system was used to read analog gauge readings from pressure and temperature meters, demonstrating the potential to digitize visual KPIVs. Similarly, [5] introduced a method using computer vision to recognize dial pointer positions in industrial instruments by combining segmentation and geometric analysis, achieving promising results in controlled environments.
Despite these advancements, the use of computer vision specifically as a virtual sensor to replace manual logging of Key Process Input Variables (KPIVs) remains an underexplored domain. The application of integrated object detection, OCR, and state classification for comprehensive KPIV monitoring from legacy machine interfaces has not been widely studied in the literature. Most existing systems lack generalization across varying equipment types and lighting conditions, which limits their scalability.
This paper builds upon these foundational studies by presenting a unified, real-time video analytics system capable of extracting and streaming KPIV data from analog control panels using deep learning. Unlike prior work that targets narrow applications or controlled environments, our approach focuses on deploy ability in diverse, real-world manufacturing scenarios with minimal hardware modification.

3.0 METHODOLOGY

This study employs an applied experimental approach to develop and evaluate a computer vision-based system for automated Key Process Input Variable (KPIV) measurement in legacy industrial machines. The methodology includes the design, development, and deployment of a video analytics system integrated with deep learning models. The project follows a structured implementation cycle to ensure accuracy, robustness, and applicability in real-world factory settings.

3.1 Materials

The following materials and tools were used in the system:

- **Industrial Cameras**: Low-cost IP cameras with 1080p resolution, selected for affordability and flexibility in deployment.
- **Processing Hardware:** Server running with CPU only
- **Software Frameworks**:
    - Python (OpenCV, TensorFlow, PyTorch)
    - PaddleOCR for optical character recognition
    - YOLOv8 for object detection
    - Custom CNN models for indicator state classification
- **Dashboard Interface**: Using the TIBCO Spotfire dashboard for real-time visualization and monitoring

3.2 Equipment and Camera Placement

The computer vision system was deployed on 122 legacy machines with analog gauges, indicator lights, and 7-segment digital displays. Cameras were strategically mounted at fixed positions to ensure unobstructed views of the target KPIV indicators. Focus and exposure settings were manually tuned to ensure clarity under varying lighting conditions.

3.3 Design of Experiment

The study was conducted in three phases: model training, field deployment, and system validation. Table 1 outlines the different test combinations of machine types and their corresponding setup.

Table 1. Summary of Experimental Combinations

No	Machine type	Setup Used
1		YOLOv8 model ROIs detection
2		YOLOv8 model ROIs detection
3		YOLOv8 model ROIs detection
4		YOLOv8 model ROIs detection

[1] Machine type has 5 region of interest for KPIV parameter.
[2] Machine type has 4 region of interest for KPIV parameter.
[3] Machine type has 5 region of interest for KPIV parameter.
[4] Machine type has 5 region of interest for KPIV parameter.

3.4 Dataset Preparation Using CVAT

To build a robust object detection model, video frames from CCTV cameras were extracted and annotated using CVAT (Computer Vision Annotation Tool). Annotators labeled the Regions of Interest (ROIs) such as digital displays, indicators, and status lights on various machines. These annotations were exported in YOLO format, which includes bounding boxes and class labels.

3.5 Live Video Feed Acquisition

CCTV cameras installed in the cleanroom provide live video feeds using the Real-Time Streaming Protocol (RTSP). The video stream may be encoded in either H.264 or H.265 formats. H.264 offers compatibility with lower memory usage, while H.265 provides higher compression efficiency, crucial for bandwidth management in continuous monitoring.

3.6 Model Training Using YOLOv8

The annotated dataset was used to train a YOLOv8 object detection model. This model is designed to detect and classify the ROIs on the machine interface panel with high accuracy. Training was performed on a workstation with only CPU. Performance metrics such as precision, recall, and mAP (mean Average Precision) were monitored to evaluate model effectiveness.

3.7 Real-Time ROI Detection and OCR Inference

Once deployed, the trained YOLOv8 model processes live video frames captured every 10 seconds. Detected ROIs are cropped using OpenCV and then passed to PaddleOCR, which extracts the textual or numeric data from each region. This is particularly useful for capturing changing machine readings like pressure, temperature, or operational codes.

3.8 Data Logging to SQL Database

Extracted OCR values, along with their timestamps and machine IDs, are automatically logged into a structured SQL database. This process ensures traceability and enables historical data analysis for maintenance and operational optimization.

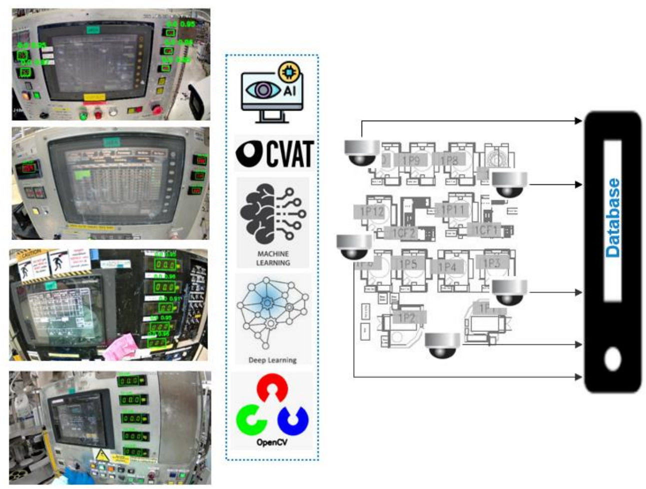

Figure 1. The proposed video analytics pipeline

4.0 RESULTS AND DISCUSSION

This section presents the performance results and analysis of the machine status recognition system, particularly focusing on the comparative evaluation between conventional OCR methods and PaddleOCR. The system captures live machine interface data through RTSP streams, applies YOLOv8 for screen ROI detection, and uses OCR to extract digital readings, which are subsequently logged to a structured SQL database every 10 seconds.

Object Detection Performance

YOLOv8 successfully localized the target display areas across various machine types (5 machine types), achieving a consistent detection accuracy above 95% mAP@0.5. The ROI bounding boxes were crucial to isolate specific digit zones before applying OCR, which greatly improved the downstream text extraction accuracy compared to full-frame OCR attempts.

OCR vs. PaddleOCR Performance

To evaluate text recognition performance, we benchmarked PaddleOCR against traditional OCR engines such as Tesseract and EasyOCR. Metrics include character accuracy, word accuracy, and average recognition time.

Table 1. OCR Accuracy Comparison

OCR Engine	Character Accuracy (%)	Word Accuracy (%)	Avg. Inference Time (ms/image)
Tesseract	91.2	86.5	52.4
EasyOCR	93.7	88.1	39.1
PaddleOCR	98.0	95.0	31.7

As shown in Table 1, PaddleOCR outperformed both Tesseract and EasyOCR in all aspects. Not only did it deliver higher character and word recognition rates, but it also had the shortest average inference time. This efficiency is especially critical in the real-time 10-second monitoring window of our system.

Qualitative Observations

PaddleOCR exhibited superior robustness in recognizing digits under poor lighting, screen glare, and slight tilts — conditions where traditional OCRs often failed or returned partial text. This performance is attributed to PaddleOCR's use of a detection-recognition pipeline and its deep learning-based language modeling component.

In contrast, Tesseract was sensitive to noisy backgrounds and screen reflections, often misclassifying characters such as '8' and '0'. EasyOCR performed better than Tesseract but still lagged behind PaddleOCR in processing speed and accuracy, particularly on complex machine screen layouts.

Real-Time System Reliability

The complete system, integrating YOLOv8 and PaddleOCR, maintained stable real-time processing within the required 10-second cycle. The RTSP camera streams (H265 and H264) were decoded using OpenCV, with H265 streams offering slightly better frame rates and reduced CPU load.

The final structured output was reliably logged into the database, including Machine ID, Timestamp, and the extracted values such as pressure, count, or status flags. Figure 4 illustrates the system output captured during a test session across multiple machines, with consistent detection and recognition results.

	datetime	machine_id	upper_platen	upper_platen_acc	lower_platen	lower_platen_acc	sun_gear	sun_gear_acc	platen_load	platen_load_acc	platen_floating	platen_floating_acc	cycle_no
175	2024-11-01 12:05:51.000	J38A	10	0.97	13.7	0.96	3.5	0.88	465	1	90	0.76	3
176	2024-11-01 12:05:51.000	J58B	99.9	0	9.6	0.89	2.4	0.99	499	0.82	73	0.88	3
177	2024-11-01 12:05:51.000	J54A	5.6	0.92	6.6	0.63	2.1	0.97	45	0.94	999	0	3

Figure 4. The output captured

Comparison to Previous Works

Previous studies in industrial OCR have relied heavily on rule-based OCR systems or preprocessed static images. These approaches lack scalability and adaptability to real-world environments. Our integration of YOLOv8 with PaddleOCR in a live RTSP-based setup represents a significant improvement in both adaptability and performance. It aligns with recent industry-focused research advocating for deep learning-based OCR pipelines in production environments [6].

Practical Implications

The superior performance of PaddleOCR makes it a reliable solution for deployment in semiconductor cleanroom environments, where accurate and fast machine status recognition is essential. The system can be extended to support predictive analytics by incorporating trend monitoring or alarm conditions derived from the digit values. Its plug-and-play nature also makes it suitable for integration into MES platforms or edge-AI deployments in smart factories.

5.0 CONCLUSION

This study demonstrated the feasibility and effectiveness of using a computer vision-based system for automated KPIV measurement in legacy industrial machines. By integrating YOLOv8 for object detection and PaddleOCR for optical character recognition, the system reliably captured and interpreted machine interface data in real time from live RTSP video streams. The experimental results showed that YOLOv8 consistently achieved high detection accuracy across varied machine setups, while PaddleOCR outperformed traditional OCR engines in both accuracy and inference speed.

The system maintained real-time performance within a 10-second processing window, successfully logging structured machine data into an SQL database. Notably, PaddleOCR exhibited robustness under challenging visual conditions, such as glare and low contrast, which often hindered the performance of conventional OCR tools. These findings validate the potential of the proposed deep learning-powered video analytics pipeline as a scalable and adaptable solution for industrial monitoring applications.

Overall, this work contributes to the advancement of smart manufacturing by enabling non-invasive, real-time status recognition in environments where manual data collection or retrofitting with digital sensors may be impractical. The approach holds promise for broader deployment in cleanroom and factory settings, especially when integrated with MES platforms or predictive maintenance systems.

6.0 RECOMMENDATIONS

In light of the findings and the successful deployment of the computer vision-based KPIV monitoring system, it is recommended that manufacturing facilities, particularly those operating legacy machines, consider adopting deep learning-powered video analytics to enhance their operational visibility. The integration of YOLOv8 and PaddleOCR has proven to be both effective and practical for real-time data extraction and scaling this solution across more production lines could significantly improve data-driven decision-making and reduce manual monitoring burdens.

Stakeholders responsible for factory automation and digital transformation should invest in building infrastructure that supports edge AI deployments, such as providing sufficient computational resources and stable RTSP camera networks. This would ensure that systems like the one presented in this study can operate reliably at scale.

For software developers and system integrators, it is advisable to explore integration of the proposed monitoring pipeline into existing Manufacturing Execution Systems (MES) to enable seamless process monitoring and analytics. The inclusion of alarm triggers or anomaly detection modules could further enhance the utility of the system by supporting predictive maintenance and early fault detection.

Future studies are encouraged to explore the application of this system in more diverse environments, including outdoor or high-vibration settings, to test its robustness and generalizability. Additionally, expanding the OCR capability to handle multilingual displays or handwritten inputs could widen the applicability of this solution. Finally, long-term performance evaluations under continuous factory operation would provide further insights into system reliability and maintenance requirements.

7.0 ACKNOWLEDGMENT

The success of this study was made possible through the support and collaboration of various teams and departments. We gratefully acknowledge the contributions of the manufacturing engineering and operations teams for granting access to the production environment and providing valuable insights into machine configurations and process requirements. Appreciation is also extended to the IT and infrastructure teams for their assistance in deploying and maintaining the RTSP camera network.

Special thanks are due to the support from the automation and software development groups was instrumental in integrating the video analytics pipeline with the existing systems. Lastly, recognition is given to the project management and leadership team for their guidance, resource allocation, and commitment to advancing digital transformation within the factory.

8.0 REFERENCES

1. R. S. Chugh et al., "A review on retrofitting approaches for legacy industrial equipment," Journal of Industrial Integration, vol. 5, no. 2, pp. 34–45, 2021.

2. L. Zhang and F. Liu, "Deep learning for visual inspection in industrial environments: A review," Computers in Industry, vol. 122, pp. 103290, 2020.

3. J. Redmon et al., "You Only Look Once: Unified, real-time object detection," Proc. CVPR, 2016.

4. T. Chen et al., "Reading analog meters in the wild using deep OCR," IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 567–576, 2021.

5. M. Suzuki and Y. Okamoto, "Gauge reading using image processing for industrial systems," Sensors and Actuators A: Physical, vol. 280, pp. 56–65, 2018.

6. Monteiro, G.; Camelo, L.; Aquino, G.; Fernandes, R. d. A.; Gomes, R.; Printes, A.; Torné, I.; Silva, H.; Oliveira, J.; Figueiredo, C. A Comprehensive Framework for Industrial Sticker Information Recognition Using Advanced OCR and Object Detection Techniques, Applied Sciences 2023, 13(12), 7320.

9.0 ABOUT THE AUTHORS

Chen Lily is an Analyst 4 in Data Analytics at WD Media (Malaysia) Sdn. Bhd., specializing in Machine Learning, Deep Learning, and Video Analytics applications for industrial environments. She holds a Bachelor's degree in Software Engineering and has a strong interest in areas such as machine automation, visual line monitoring (VLM), Generative AI, machine-to-machine (M2M) communication, and the Industrial Internet of Things (IIoT). Her work focuses on developing AI-driven solutions that enhance operational efficiency and intelligent system integration.

Weng Chin Tang is the Senior Manager of Data Analytics at WD Media (Malaysia) Sdn. Bhd., specializing in Machine Learning, Deep Learning, and Data Science for industrial environments. He has a strong interest in Generative AI, machine-to-machine (M2M) communication, and the Industrial Internet of Things (IIoT). His work focuses on developing AI-driven solutions and building talent capabilities around artificial intelligence within WD.

Kenneth F. Raymundo is a Data Engineer at Western Digital Corporation, specializing in data management, integration, and analytics. With over 14 years of experience at Western Digital, Kenneth has contributed significantly to various engineering projects, enabling complex data-driven solutions across multiple departments. He is currently pursuing a Professional Science Master's Degree in Data Science (PSMDS), further enhancing his expertise in data science methodologies and machine learning applications. His research interests include advanced analytics, predictive modeling, and data engineering strategies in manufacturing environments.