Video Analysis Solution for Workplace Task Recognition

Alongside the advances in manufacturing innovation that are happening around the world, Hitachi has developed video analysis technology that uses AI for the real-time recognition of the actions and dynamic characteristics of people from large amounts of video data. This technology has been introduced to the manufacturing workplaces. Suitable for a variety of applications, this technology can be used for the remote and non-contact recognition of people's actions at various different levels of granularity, such as where they are on the factory floor and what paths they follow, how they perform tasks such as assembly work, or their skill at the sort of fine-scale work that is done using a microscope. Work quality, productivity, and safety in manufacturing can all be improved by augmenting 4M data with the results of this analysis of people's factory floor actions.

Hiroshi Yoshikawa
Shinya Kaneko
Takehiro Urano
Hiroto Nagayoshi
Toshihiro Ohta

1. Introduction

Innovations in manufacturing are happening around the world, with the concept of Industrie 4.0 leading the way. This is aimed at bringing together the various different elements such as people, goods, systems, and companies to create value in order first to provide transparency in the manufacturing workplace and ultimately to improve quality and productivity across the entire supply chain.

Hitachi believes that the pathway to manufacturing innovation lies in the ongoing execution of the sense-think-act process that involves the sensing of "4M data" from manufacturing workplaces (meaning data that relates to humans, machines, materials, and methods), linking it together for analysis, and then carrying out the actual instructions or other control operations. Hitachi has been focusing in particular on people in the workplace, an area from which it has been difficult to collect quantitative data in the past. This involves developing and supplying technology in which video analysis is used for things like the detection of deviations from correct practice during assembly work or the capturing of expert skills in digital form(1), (2).

Maintaining business continuity under conditions where work needs to be non-contact or done remotely has become an issue in recent years due to frequent natural disasters and the spread of COVID-19. At manufacturing workplaces, along with advances in automation, there is also an increasing need for remote, non-contact monitoring and support of the manual work that still remains despite this automation. Meanwhile, advances in 5th-generation (5G) networking and edge computing have made it easier to analyze large amounts of video data from manufacturing workplaces. This has raised hopes that video analysis can be used to improve productivity, quality, and safety in manufacturing.

2. Action Recognition at Manufacturing Workplaces

2. 1 Categories of Solutions for Action Recognition

Hitachi has responded to this demand from the manufacturing industry by supplying solutions that can recognize worker actions at manufacturing workplaces. In reality, the term “worker action” covers a wide range of movements that depend on the particular materials (what is being produced) and methods (how it is being produced). Accordingly, Hitachi’s solutions offer a range of different action recognition options to suit the granularity with which these actions are to be recognized and the relevant feature values. This action “granularity” can be divided into the following three categories.

(1) Location and path traveled

Tracking where a person is on the factory floor.

(2) Posture and movements

Tracking a person’s posture and their body movements.

(3) Fine-scale actions

Tracking a person’s body movements when doing fine-scale work (such as work on precision parts)

The following are examples of the sort of feature values that can be obtained depending on what action recognition is being used for.

(1) Productivity improvement: Work duration, start, and end times

(2) Quality improvement: Whether there are any deviations from standard work practices, how tools are used (tool position, speed, and so on)

(3) Safety improvement: Whether anyone enters restricted areas, frequency of adopting an unsafe posture

The following sections describe solution use cases that involve different action granularities.

2. 2 Detection of Location and Path Traveled

This use case involves measuring actual task duration in the assembly work area without requiring the worker to take any special action. As shown in **Figure 1**, workers assembling products in a work area use voice control to access three-dimensional (3D) work procedure manuals. While it is possible to determine how long each task takes from logs of worker interaction with the procedure manuals, the problem is that accurate task duration measurements cannot be obtained in cases such as when a worker leaves the work area for some reason in the middle of a task.

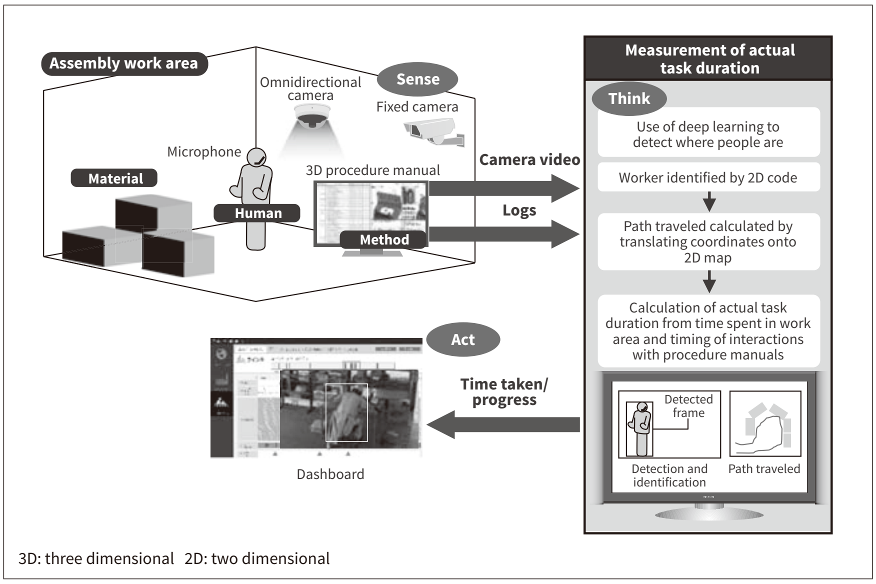
To overcome this challenge, a number of cameras are installed in the work area (sense) and the actual time spent on assembly work is measured by combining the results of using video analysis to detect where people (human) are with logs (method) of their interaction with procedure manuals. Determining where people are is done by combining use of deep learning to detect their presence and use of two-dimensional (2D) barcodes to identify them. A deep learning technique is used to detect people in video images and their identity is then determined by looking for a 2D barcode in their region of the image. Based on this, the task duration is calculated from the logs and the total time spent in the work area (the extent of which is predefined).

To verify how well the technique works, it was used to detect people’s locations while performing actual tasks at a workplace, using this to measure the duration of these tasks. The results demonstrated that actual work times can be measured with a mean error of 3%.

Along with this use for improving productivity by measuring how long it takes workers to complete tasks in the

Figure 1 — Example of Solution that Detects People’s Location

This solution measures the time taken to complete work by combining the detection of people’s location with logs of worker interaction with procedure manuals.


Assembly work area
Omnidirectional camera
Sense
Fixed camera
Microphone
3D procedure manual
Material
Human
Method
Camera video
Logs
Measurement of actual task duration
Think
Use of deep learning to detect where people are
Worker identified by 2D code
Path traveled calculated by translating coordinates onto 2D map
Calculation of actual task duration from time spent in work area and timing of interactions with procedure manuals
Detected frame
Detection and identification
Path traveled
Act
Time taken/ progress
Dashboard
3D: three dimensional 2D: two dimensional

case of cell production, or how long it takes supervisors and parts suppliers to complete tasks in the case of line production, the ability to detect where people are and track their movements can also be used to improve safety, such as by managing access to dangerous areas.

2. 3 Recognition of Posture and Movements

This use case involves measuring the time taken for each step on a production line without requiring workers or supervisors to take any special action. Past practice was to analyze signals from the conveyors (machine) that transport products along the line to calculate task durations from the times of product arrivals and departures at each station. On production lines where conveyor progress is synchronized across all stations, however, workers may sometimes find themselves waiting for upstream or downstream work to finish. In such cases, the accurate measurement of actual task durations requires that workers record their task start and end times or that a supervisor measures them visually.

To overcome this, cameras are installed where they can observe workers at each station on the line and recognize what they are doing based on how their hand positions or other body positions change over time. By this means, it is possible to distinguish between the posture of someone working on a product on the conveyor and someone who is waiting because they have nothing to work on.

Figure 2 shows an overview of the algorithm for measuring actual task durations. Using frames from the work video as input, the algorithm uses deep learning to perform posture recognition in a way that identifies the positions of the worker's joints. The sequence of feature values calculated from the worker joint positions obtained by posture recognition are then used as the input to task recognition, which also uses deep learning to identify what is being done and distinguish between times when the worker is performing a task and times when they are idle. A neural network has been developed for task recognition that is based on long short-term memory (LSTM), a proven technique for the analysis of time-series data. The results of task identification (human) are then combined with the signals from the conveyor (machine) to calculate the actual task duration.

To verify how well the developed technology works, it was used to measure the durations of actual tasks at a workplace. The results demonstrated that, for assembly work with a takt time of 180 s, actual task durations can be measured with an error of less than ±3 s by distinguishing between times when workers are performing a task and when they are idle. Further analysis of the measured task durations could be used to improve productivity, such as by improving how particular tasks are performed or by optimizing line balance.

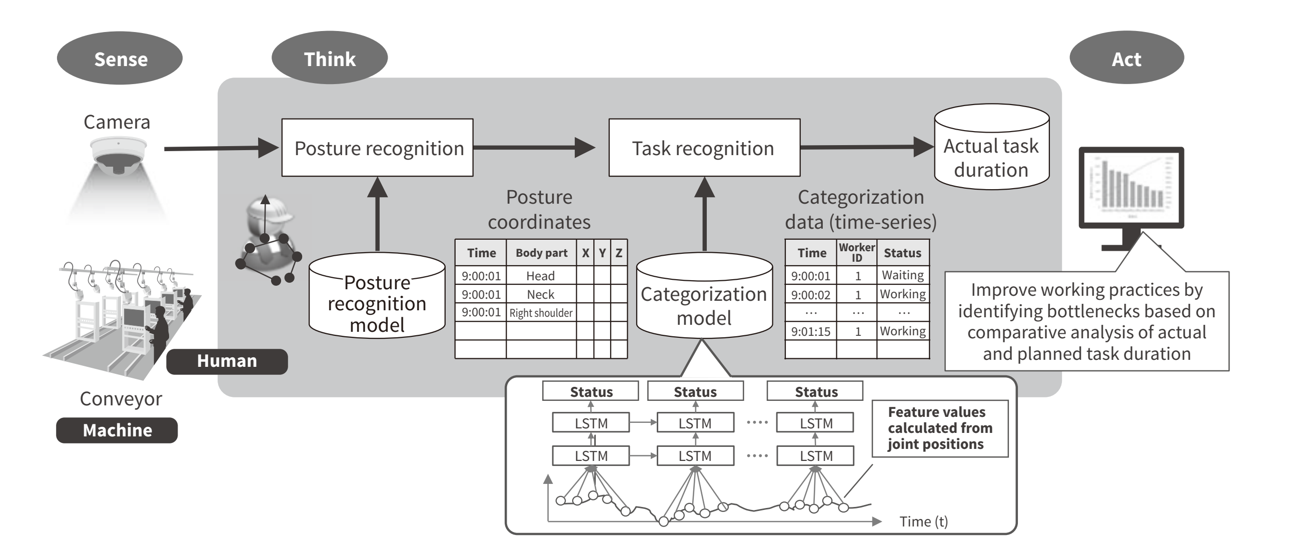
Along with its above use for measuring the actual durations of each task performed on a production line, the recognition of posture and movement can also be used for things like detecting deviations from standard work practices (quality improvement) or when workers are in danger (safety).

2. 4 Recognition of Fine-scale Actions

This use case applies to the sort of fine-scale work that is done using a microscope and requires 3D positioning with sub-millimeter (0.1–1 mm range) accuracy. Training workers to be proficient in the delicate task of micromachining precision equipment has in the past mainly been based

Figure 2 — Measurement of Actual Task Duration Using Posture Recognition

Actual task duration is calculated by distinguishing between times when the worker is performing a task and times when they are idle, which is done based on the changes over time in the results of worker posture recognition.


Sense
Think
Act
Camera
Posture recognition
Task recognition
Actual task duration
Posture recognition model
Posture coordinates
Categorization model
Categorization data (time-series)
Human
Conveyor
Machine
Improve working practices by identifying bottlenecks based on comparative analysis of actual and planned task duration
Status
Status
Status
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
Feature values calculated from joint positions
Time (t)

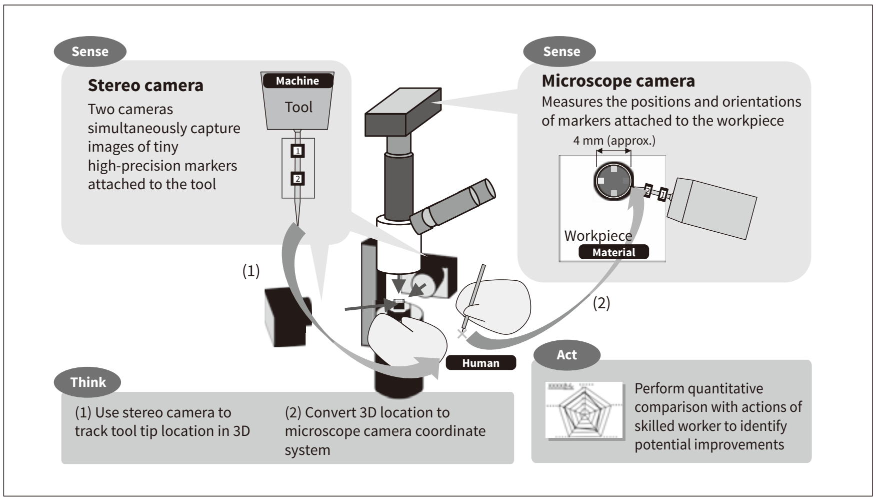
Time	Body part	X	Y	Z
9:00:01	Head			
9:00:01	Neck			
9:00:01	Right shoulder			

Time	Worker ID	Status
9:00:01	1	Waiting
9:00:02	1	Working
…	…	…
9:01:15	1	Working

LSTM: long short-term memory ID: identification

Figure 3 — Fine-scale Action Recognition System

The system can determine tool movements in three dimensions with sub-millimeter accuracy by using a stereo camera that records images of tiny high-precision markers attached to the tool being used and a microscope camera that records the microscope field of view.


Sense
Stereo camera
Two cameras simultaneously capture images of tiny high-precision markers attached to the tool
Machine
Tool
1
2
Sense
Microscope camera
Measures the positions and orientations of markers attached to the workpiece
4 mm (approx.)
Workpiece
Material
(1)
(2)
Human
Act
Think
(1) Use stereo camera to track tool tip location in 3D
(2) Convert 3D location to microscope camera coordinate system
Perform quantitative comparison with actions of skilled worker to identify potential improvements

on qualitative instruction, whereby the skills are acquired through a process of trial and error in which the quality of work they deliver is repeatedly reviewed. As a result, it takes several months to train someone to the required level.

In response, Hitachi has developed a system for recognizing fine-scale actions that converts these into digital form to enable the skills of regular workers to be compared with those of highly skilled ones. The aims are to speed up the training of skilled workers and to raise the standard of workmanship.

Figure 3 shows a diagram of how the recognition system works. Two cameras are positioned next to the worker (forming a stereo camera) and both record images of two tiny high-precision markers that are attached to the tool used by the worker. Another camera (a microscope camera) is attached to the lens barrel of the microscope used by the worker and records images of what they see as they work. The object they are working on (the workpiece) also has a number of markers attached to it that are captured in these images.

The stereo camera measures the 3D positions of the two markers on the tool from which the position of the tool tip can be inferred. This position is then translated into the coordinate system of the microscope camera based on the geometric relationship between the stereo camera and microscope camera images that has been determined in an earlier calibration step. The microscope camera is then used to measure the workpiece marker locations from which the movements of the tool relative to the workpiece can be obtained with sub-millimeter precision.

To verify how well the technology works, it was used to track fine-scale actions during actual work at a workplace. The results demonstrated that actions performed in a work area of approximately 4 mm (specifically, the position of the tool tip) could be determined with sub-millimeter accuracy. The intention is to deploy the system in the future for uses such as comparing the actions of regular workers with those of highly skilled ones and speeding up the training of skilled workers by analyzing their correlations with work quality.

3. Future Outlook

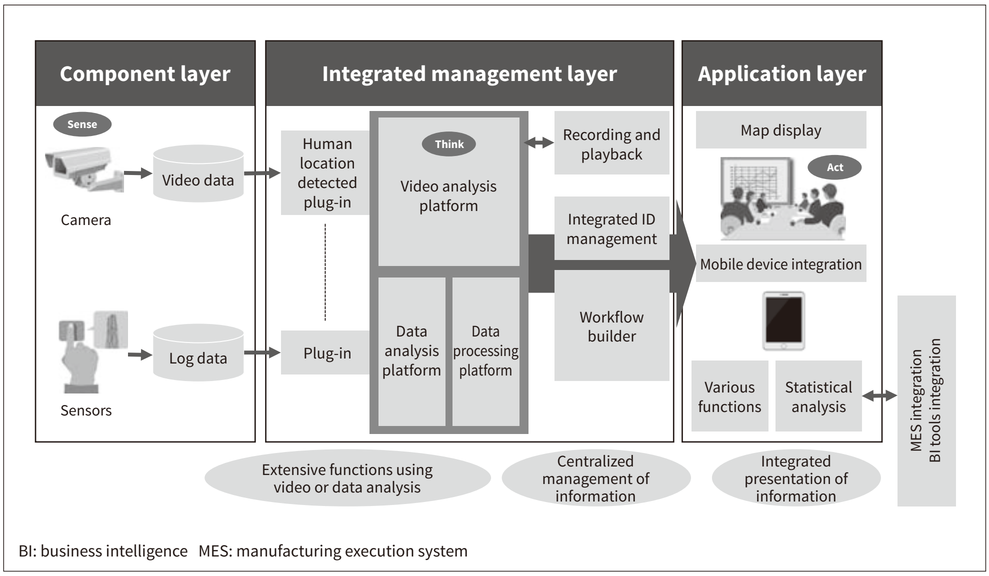
Rapid advances in deep learning over recent years have seen a steady stream of new video analysis technologies with potential for use in action recognition. Wearable devices equipped with inertial or other types of sensors have also become available in a variety of forms, such as those that can be worn like a shirt. Achieving a rapid resolution of the challenges facing manufacturing workplaces will require that these technologies be combined into systems in appropriate ways.

Hitachi Industry & Control Solutions, Ltd. has developed and deployed an integrated platform for physical security that features the coordinated collection, archiving, and analysis of video data and various other forms of Internet of Things (IoT) data[3]. The platform can be used to get video analysis systems up and running quickly, with a variety of video analysis technologies available for use as modular plug-ins (see **Figure 4**). Through interoperation between this platform and other systems, Hitachi intends to link human data from the analysis of people with 4M data from the workplace and to put this to use in resolving the challenges facing manufacturing workplaces.

Meanwhile, the infrastructure for video analysis is becoming more readily available, with advances in networking, especially 5G, making it easier to set up factory networks with high speed and wide coverage. Unfortunately, the cost of collecting, archiving, and analyzing large amounts of video data on a single platform remains prohibitive. This is creating a need for ways of performing video analysis in real time at locations close to the workplace, or more specifically close to the cameras from which the data is generated (what is called edge computing). Accordingly, Hitachi intends to put in place the technologies that will enable the rapid and low-cost delivery of solutions that take advantage of this edge computing and 5G networking.

Figure 4 — Overview of Integrated Platform for Physical Security

By offering a variety of video analysis technologies as plug-ins, the platform enables video analysis systems to be configured quickly.



BI: business intelligence MES: manufacturing execution system

Furthermore, Hitachi also aims to optimize production and improve quality across entire supply chains by using its digital twin solution[4] to link together workplace 4M data collected by these means across the different activities and organizations that make up the supply chain.

4. Conclusions

This article has described Hitachi's work on action recognition that uses video analysis of where people are and where they go, of posture and movement, and of fine-scale actions to improve productivity, quality, and safety at manufacturing workplaces. In the future, Hitachi intends to utilize techniques for the integration of edge and cloud computing to deliver solutions quickly and at low cost.

Acknowledgement

Considerable assistance was received from Daikin Industries, Ltd. and other interested parties in the development of the video analysis system described in this article. The authors would like to take this opportunity to express their deepest gratitude.

References

1) H. Isaka et al., "Next Generation of Global Production Management Using Sensing and Analysis Technology," Hitachi Review, Vol. 65, No. 5, pp. 47–52 (Jun. 2016).

2) H. Umeki et al., "Digital Encapsulation of Manufacturing Site Expertise Utilizing Sensing Technology," Hitachi Review, 67, pp. 199–205 (Feb. 2018).

3) H. Okita et al., "AI-based Video Analysis Solution for Creating Safe and Secure Society," Hitachi Review, 69, pp. 687–693 (Sep. 2020).

4) D. Ito et al., "Digital Twin Technology for Continuous Improvement at Manufacturing Sites," Hitachi Review, 69, pp. 644–648 (Sep. 2020).

Authors

Hiroshi Yoshikawa
Industrial FA Solution Department, Industrial Manufacturing Solution Division, Industrial Solutions & Services Business Division, Digital Solutions Division, Industry & Distribution Business Unit, Hitachi, Ltd. *Current work and research:* Development and delivery of digital solutions for industrial factories. *Society memberships*: The Information Processing Society of Japan (IPSJ).

Shinya Kaneko
Industrial Visual Solution Department, Industrial Manufacturing Solution Division, Industrial Solutions & Services Business Division, Digital Solutions Division, Industry & Distribution Business Unit, Hitachi, Ltd. *Current work and research:* Development and delivery of visual solutions for industrial factories.

Takehiro Urano
Industrial FA Solution Department, Industrial Manufacturing Solution Division, Industrial Solutions & Services Business Division, Digital Solutions Division, Industry & Distribution Business Unit, Hitachi, Ltd. *Current work and research:* Development and delivery of digital solutions for industrial factories. *Society memberships*: The Society of Instrument and Control Engineers (SICE).

Hiroto Nagayoshi
Center for Technology Innovation – Artificial Intelligence, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development regarding computer vision and applications. *Society memberships*: IPSJ and the Institute of Electronics, Information and Communication Engineers (IEICE).

Toshihiro Ohta
Physical Security Solution Design Department, Social & Public Infrastructure Systems Division, Hitachi Industry & Control Solutions, Ltd. *Current work and research:* Development and delivery of visual solutions for industrial factories.