Feature Extraction System during Bolt Tightening Work by Regression CNN and Classification CNN

KOICHI MURAI
Nagasaki University
Nagasaki, Japan
murai@sys-inoue.co.jp

TETSUO IMAI
Nagasaki University
Nagasaki, Japan
imai@nagasaki-u.ac.jp

KENICHI ARAI
Nagasaki University
Nagasaki, Japan
k.arai@cis.nagasaki-u.ac.jp

TORU KOBAYASHI
Nagasaki University
Nagasaki, Japan
toru@cis.nagasaki-u.ac.jp

***Abstract*— This paper proposes a general-purpose system to extract motion features of workers during an operation at manufacturing sites from video footage. Currently, there are several systems that are used to extract motion for specific types of work. However, most of these systems are difficult to apply across many types of work. In order to make our system applicable to a variety of work at manufacturing sites, our system uses end-to-end deep learning models to extract motion features. We have applied our system to bolt tightening work using a spanner and have confirmed the effectiveness of our method.**

***Keywords*— deep learning, motion feature extraction, manufacturing**

I. INTRODUCTION

At manufacturing sites, continuous improvements in productivity are required. In particular, it is important to extract and analyze features from human motions to eliminate wasteful motions and reveal valuable ones. At the moment, several methods exist for extracting human motions, such as wearable sensors or image processing. However, most of the existing methods are limited to specific manufacturing sites or designated motions, and are unsuitable for general use. Therefore, we propose a generic method to extract motion features using two kinds of Convolutional Neural Networks (CNN). One is regression CNN for estimation of object locations, and the other is classification CNN for the classification of motions. In this paper, we have been able to use these methods to distinguish between valuable motions and wasteful motions in bolt tightening work.

II. RELATED WORK

Motion measurement and extraction are frequently used to help workers improve productivity, and several methods have been proposed.

For example, the measurement of workers' motion using a three-axis accelerometer and an unsupervised measurement method has been proposed for the estimation of lead-time for workers in assembly tasks [1]. The goal of this method was to find a start time and end time for the operation and did not focus on an analysis of problems with the operation process. Our study aims to extract motion features in order to distinguish valuable motions from wasteful motions during an operation.

In order to analyze the productivity of workers, Kitazawa proposed the method of using an acceleration sensor and beacon [2]. This method analyzes workers' positions using the beacon's received signal strength and interprets whether workers' movements are normal or abnormal from acceleration sensor data. Although it is important to find out whether workers' motions are normal or abnormal, our approach aims to distinguish between valuable motions and wasteful motions of the worker. Even though an acceleration sensor is able to collect information at high speed, there is a concern that attaching sensors to workers might interfere with their work. For this reason, we have decided to use video footage from a web-camera.

In the field of surgery, the method of CNN using images has been proposed to create an automated benchmark for surgical skills[3]. A study conducted by Zia used Space-Time Interest Points (STIPs) [4] to extract motion information from video footage. However, STIPs make it difficult to track each part of the object when analyzing motions. Our method utilizes CNN to estimate the location of objects so that we can track part of the object.

III. FEATURE EXTRACTION SYSTEM

A. Use Case

Our system extracts motion features for analysis in order to improve productivity at manufacturing sites. We focused on the motion of bolt tightening as one of the general works by a human (Fig. 1). The type of bolt tightening motion in human work is categorized into six classes based on expert motions by our criterion.

- Interval (INT): Interval among bolt tightening
- Large Movement (LM): Large movement of the spanner to the bolt tightening position
- Medium Movement (MM): Medium movement of the spanner to the bolt tightening position
- Small Movement (SM): Small movement of the spanner to the bolt tightening position
- Slow Tightening (ST): Bolt tightening at slow speed
- Fast Tightening (FT): Bolt tightening at fast speed

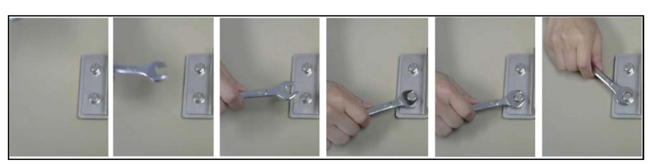

Fig. 1. Bolt tightening work

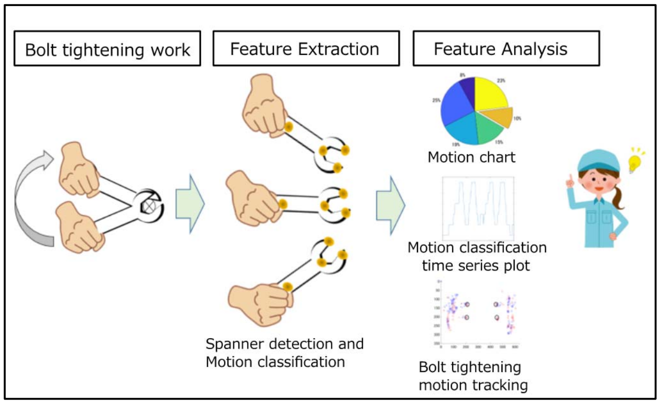

Fig. 2. Use case of the feature extraction system

Our system extracts motion features such as the locations and movement distances of each estimation point in the spanner. The motion features of expert workers and novice workers are compared in order to distinguish which motions are valuable and which are wasteful (Fig. 2). In general, an expert's motions are highly efficient and therefore valuable for productivity. Conversely, a novice's motions usually include wasteful movements. Therefore, comparing the motion features of experts and novices can reveal valuable motions and wasteful ones.

B. System Configuration

Our motion feature extraction system is composed of regression CNN and classification CNN (Fig. 3).

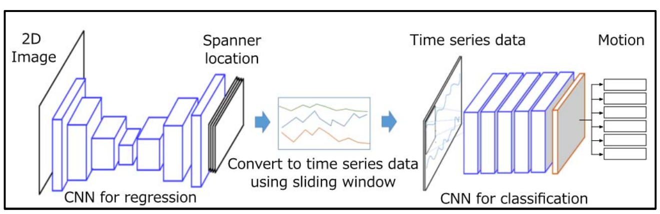

Fig. 3. CNN models

- Estimation of object locations from an image

  First, when extracting motion features it is necessary to locate the object. Our model for spanner location estimation from an image uses regression CNN based on the ResNet-18 [5] (Fig. 4). Our model has added three de-convolutional layers, a convolutional layer, and a regression layer to ResNet-18 to generate predicted heat maps for estimation of each location at four detection points (Left top, Right top, Center, Handle) in the spanner (Fig. 5). This simple approach is the same one used to estimate human poses [6].

- Classification of motions during work

  Our classification CNN is a simple Fully Convolutional Network (FCN) which is composed of five convolutional layers and classifies time series data into six motions in the same way as the image classification (Fig. 6). Time series data are multivariate data consisting of eight dimensions for locations of detection points in the spanner and four dimensions for each distance traveled. By using a sliding window method, we are able to use this information as time series data. In the experiment, we have used 10 frames for the window's size, and the frame rate used is 10 frames per second.

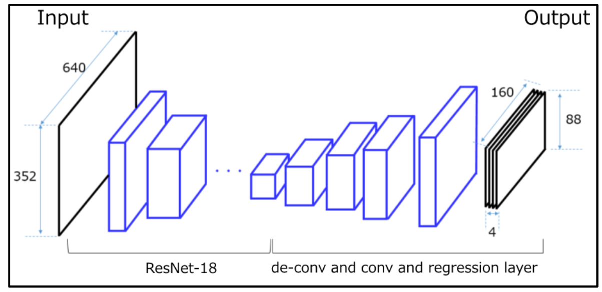

Fig. 4. CNN for estimation of object locations

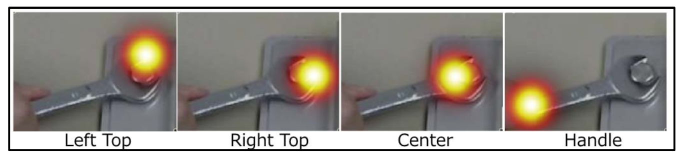

Fig. 5. Detection points in the spanner

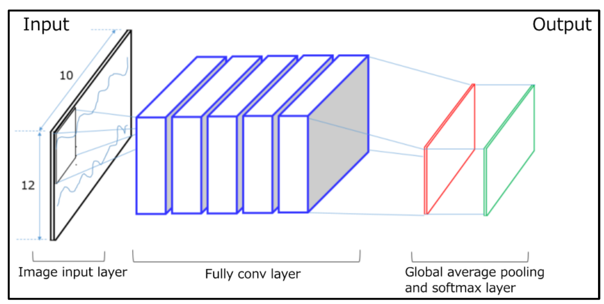

Fig. 6. CNN for classification of motions

C. Training and Evaluation

- Training of models

  We trained our networks as shown in Table I and an expert dataset was used for the training of each network. The results of the training are shown in the learning curves (Fig. 7, 8).

TABLE I. TRAINING PARAMETER

CNN	**Training parameters**			
	Solver	***Number of training Data***	***Mini batch***	***Epoch***
Regression CNN	Adam	500	32	100
Classification CNN	Adam	4,000	128	500

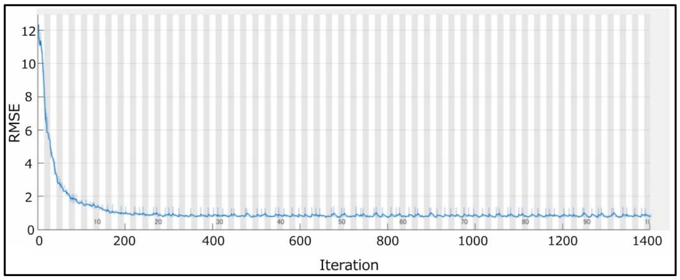

Fig. 7. learning curve of regression CNN

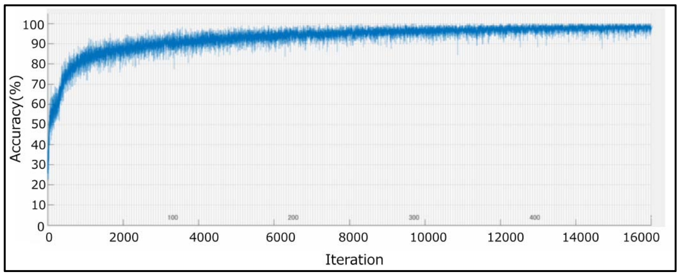

Fig. 8. learning curve of classification CNN

- Accuracy of motion classification

We evaluated the proposed system through an experiment that used three workers from our company (Expert, Novice A, Novice B) as subjects. Regarding valuation accuracies, we used an expert dataset and a novice dataset that were not used for the training of the network. The average accuracies of motion classification were 94.31% by the Expert, 93.41% by the Novice A subject, and 93.58% by the Novice B subject (Table Ⅱ).

TABLE II. MOTION CLASSIFICATION ACCURACY

Worker	**Motion Classification accuracy (%)**						
	INT	***LM***	***MM***	***SM***	***ST***	***FT***	***Ave.***
Expert	100	94.44	94.74	91.49	94.70	90.48	94.31
Novice A	100	96.55	86.36	91.67	92.11	93.75	93.41
Novice B	100	93.88	89.47	92.68	96.15	89.29	93.58

D. *Feature Analysis*

We show the results of the analysis by using motion classifications as features of each worker. The conditions of this experiment were the following.

- Three subjects participated in the experiment. (Expert, Novice A, and Novice B)
- They each conducted bolt tightening work twice.
- Internal or external factors were eliminated.

- Percentage of each motion class

We understand that there is a difference in the ratio of movements between the Expert, Novice A, and Novice B (Fig. 9). Novices use a higher proportion of the slow tightening (ST) motion compared to the Expert.

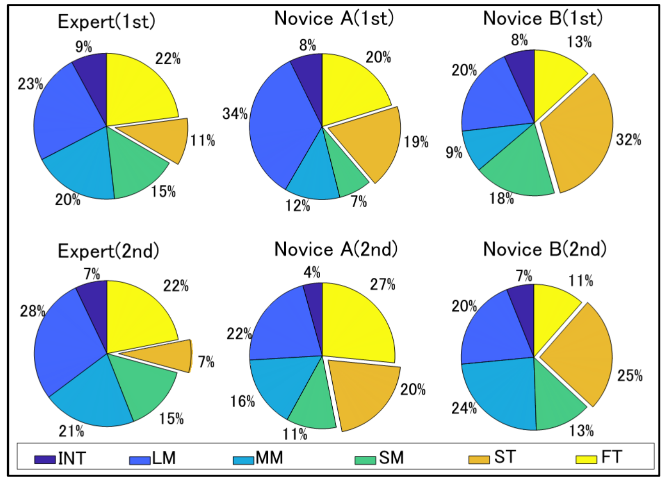

Fig. 9. Motion chart

- Variations in each motion

The variations in each motion are plotted on time series graphs in Figure 10. These graphs show motion classification results during two cycles of bolt tightening work. The Expert repeats similar motions for the bolt tightening operation. However, Novice B's motions vary with each operation.

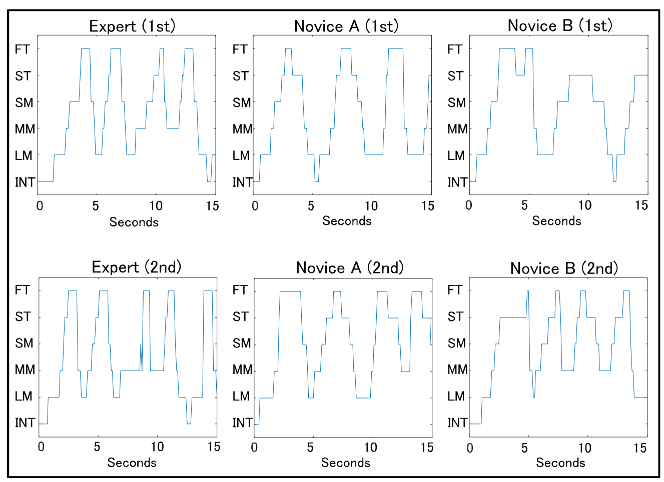

Fig. 10. Motion classification time series plot

- Movement range of the spanner

Figure 11 shows the locations of the center and handle of the spanner during the bolt tightening operation at four bolting points. This figure makes it easier to understand the four bolting positions for the motion area of the spanner. In general, it is more efficient to rotate the spanner at an angle of 60 degrees at once (Fig. 12). Figure 11 indicates that the Expert uses a smooth rotation with an angle of 60 degrees. However, Novice B's motion is slower, and movements were limited to a narrow range.

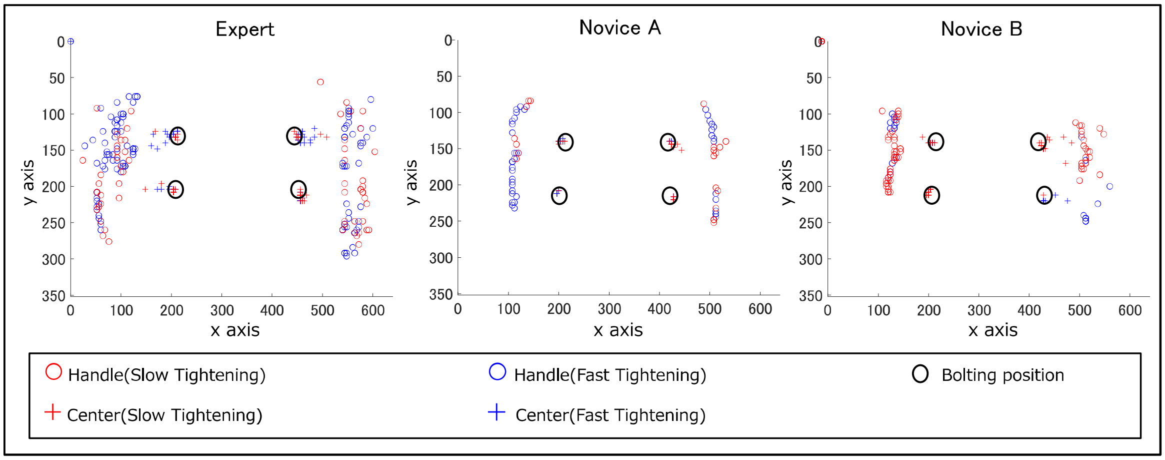

Fig. 11. Bolt tightening operation tracking

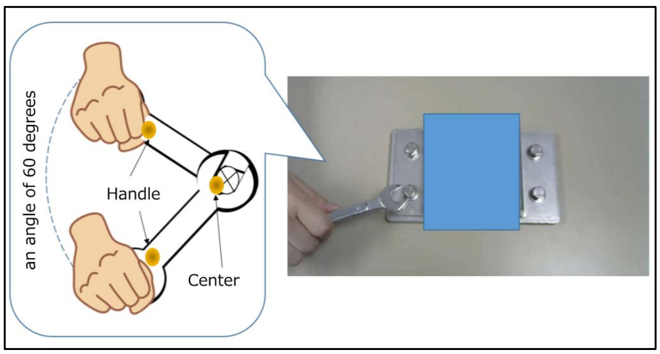

Fig. 12. Spanner rotation angle

IV. Conclusion

This paper demonstrates that our system can extract the motion features of workers using two types of CNN. As a result of our evaluation of extracted motion features, we were able to distinguish between valuable motions and wasteful motions in bolt tightening work.

In the future, we will continue to evaluate the applicability of our system in two ways. First, our system will be evaluated on other types of work such as hammering and welding. These additional tests will confirm our system's versatility. Second, we will test whether our system can detect differences in work efficiency due to external pressures or fatigue. As a result, this system might help determine drop-offs in work efficiency due to external or internal factors and prevent work mistakes.

Further, we will compare our system with other specialized approaches that use various sensors. For example, an acceleration sensor is able to accurately estimate motion speed, and a gyro sensor is able to accurately estimate an angle of motion. However, these sensors might have difficulty locating the object. Although our approach is less accurate than sensors, it is able to extract comprehensive features, including the location, speed, and angle using only a web-camera. We will identify the strengths and weaknesses of each approach through further study.

References

[1] Yasuo Namioka, Daisuke Nakai, Kazuya Ohara, Takuya Maekawa, "Automatic Measurement of Lead Time of Repetitive Assembly Work in a Factory Using a Wearable Sensor," Journal of Information Processing, vol. 25, pp. 901–911, Oct. 2017.

[2] Masaki Kitazawa, Satoshi Takahashi, Toru B. Takahashi, Atsushi Yoshikawa, Takao Terano, "Real Time Worker's Behavior Analyzing System for Productivity Measurement Using Warable Sensor," SICE Jurnal of Control, Measurement, and Ssystem Integration, Vol. 10, No. 6, pp. 536-543, Nov. 2017.

[3] Aneeq Zia, "AUTOMATED BENCHMARKING OF SURGICAL SKILLS USING MACHINE LEARNING," Georgia Institute, Dec. 2018.

[4] Ivan Laptev, Tony Lindeberg, "Space-time interest points," IEEE International Conference on Computer Vision, pp. 432-439, 2003.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[6] Bin Xiao, Haiping Wu, Yichen Wei,"Simple Baselines for Human Pose Estimation and Tracking," 15th European Conference on Computer Vision, 2018.