Development of a Workers' Behavior Estimation System Using Sensing Data and Machine Learning

R. Tanaka[a], N. Yabuki[a], and T. Fukuda[a]

[a]Department of Sustainable Energy and Environmental Engineering, Osaka University, Japan
E-mail: rikuto.tanaka@it.see.eng.osaka-u.ac.jp, yabuki@see.eng.osaka-u.ac.jp, fukuda@see.eng.osaka-u.ac.jp

Abstract –

Accurate information on workers' behavior is important for safety and productivity management on construction sites. In recent years, some methods for estimating construction workers' behavior using sensing data have been proposed to collect the data based on scientific evidence. Due to the limitations of previously proposed methods that usually relied on expensive devices such as motion capture systems, the huge amount of investment on the system installation and human resource costs are required. This paper proposes a method for estimating workers' posture with Long Short-Term Memory (LSTM) by using terminals that have already been introduced to construction sites, taking into consideration the operational cost and issues in the previous studies. Moreover, we also propose and evaluate a data augmentation method for utilizing limited training data sets. The experiment results for a reinforcing bar worker indicated that the proposed method could estimate not only the forward-leaning and squatting postures with 79% or more F-measure but also the number of rebar binding points by the acceleration data. Besides, we confirmed that the data augmentation method improved the accuracy of posture estimation by 5%.

Keywords –

Construction worker; Sensing data analysis; Machine Learning; Behavior estimation

1 Introduction

While the number of labor accidents at construction sites in Japan is decreasing, the number of serious accidents continues to be higher than in other industries [1]. Therefore, it is necessary to take improvement measures in consideration of both the organization and the technology through a meeting among construction-related people. Moreover, as construction demand is expected to be stabilized in recent years, there is a concern that the labor force will be insufficient due to a decrease in the number of young employees and the aging of skilled technicians. Thus, regarding labor productivity as well, it is necessary to formulate efficient construction plans by saving labor costs while also considering the safety of workers [2].

In recent years, for managing construction plans that consider safety and efficiency at construction sites, the introduction of construction support systems that have the function of visualizing the situation of workers and equipment on the site from past accumulated data has been promoted [3]. At such time, it is important to collect the condition data of the workers by using informative equipment such as sensors and cameras and convert them into information that is practical for site management such as the behavior history of the workers.

Video data acquired by RGB cameras and sensing data from wearable terminals are used to gather behavior data of construction workers. In this study, we employ a method using sensing data that enables data acquisition in consideration of the obstructions on the site and personal privacy and estimates the workers' behavior [4]. Also, it has been reported that supervised learning can be used for behavioral estimation based on individual characteristics.[5]. By applying these methods to workers at construction sites, the estimation of tools used by workers [6] and the work estimation of reinforcing bar workers [7] have been performed. Furthermore, smartphones with built-in inertial sensors or motion capture systems that attached multiple inertial sensors to the joints of the workers were also proposed to estimate their behavior [8]. However, these equipment costs are not easily achievable and besides, it is also essential to prepare the human resources for security and maintenance of these devices. Therefore, it is necessary to consider constraints such as project size and budget when applying it to the field.

In this study, thus, we propose a construction worker behavior estimation system that considers the limitations when introducing it to a construction site by using monitoring devices that have already been used at the construction sites. Our method considers constraints such as sampling frequency when collecting data by adopting a model corresponding to time series data as a behavior estimation algorithm. Besides, we propose a data augmentation method because it is labor-intensive and costly in order to prepare a high amount of training data for the behavior estimation model. Finally, we verify the posture estimation accuracy of the proposed system by conducting a verification experiment using the sensing data of the workers acquired at the actual construction site.

2 Literature review

Human behavior estimation technology has gained attention in recent years such as in medicine and engineering scopes [9]. Data collection for human behavior estimation can be separated into two types: a vision-based method, in which video devices are used to collect data from a target person at a distance, and a sensor-based method, which utilizes a device with a built-in sensor directly attached to the target person. In this section, we summarize the applications of each method that have been implemented on construction sites.

2.1 Vision-based methods

Video data is used in various fields because it is easy for humans to intuitively understand and obtain useful information from the images by looking at them directly [10]. In recent years, due to cost reduction, downsizing, and high resolution of video equipment, it has been introduced to the construction sites as well, and research on data collection of worker status and construction machines' positions have been conducted [11]. In particular, recent research has been carried out to apply computer vision technology using deep learning. Some of the studies were to automatically identify which task a worker is engaged in [12] and to verify the appropriate utilization of safety devices such as safety belts [13]. In those studies [12-13], the practitioners use cameras to track the workers. However, hiring extra staff to work on these tasks increases the cost of the project. A fixed-point camera-based method [14] has also been proposed. However, on the construction sites where workers and equipment are densely packed, the target workers may be hidden behind the equipment and the data cannot be continuously collected.

In recent years, a method using a depth camera for behavior estimation[15] has also been reported. Depth information makes it possible to correctly reproduce the human posture that occurs in the real space. However, the depth camera cannot accurately obtain depth information of distant objects and may face some difficulties when some objects are exposed to sunlight.

2.2 Sensor-based methods

The sensor-based method is proposed to eliminate the weaknesses of behavior estimation using video data. The posture and motion can be estimated by attaching a terminal with built-in inertial or biometric sensors to the human body and performing the calculation on the sensing data. This method can compensate for the shortcomings of vision-based method because it can continuously collect data without being affected by surrounding obstacles, light sources, and sight distance [4]. Also, sensor-based method can collect data in consideration of the privacy of the subject [4]. Behavior estimation using sensing data has been applied in a variety of fields, as the recent spread of micro-electro-mechanical-systems (MEMS) technology made it possible to easily develop behavior estimation systems [9]. In the construction sector, a motion capture system using multiple inertial sensors attached to human body to prevent Work-Related Musculoskeletal Disorders (WRMDs) was proposed at a construction site, and verification experiments have shown the effectiveness of the warning function [8]. However, the motion capture systems are hardly affordable and attaching many sensors to the human body is intrusive. With respect to productivity management, smartphone-based methods have been proposed to estimate tools handled by workers [6] and work estimation for rebar workers [7]. It is cost-effective to use personal smartphones for data collection which are ubiquitous these days. Nevertheless, the recent diversification of smartphone models and specifications leads to an increase in the burden for system administrators. Thus low-cost, nonintrusive and uniform equipment is suitable for installation to construction sites.

2.3 Objective of this research

Considering the constraint for installing the equipment as shown through the literature review, we adopt the monitoring devices for the workers that have already been installed to construction sites. The devices meet the requirements of low-cost, nonintrusive and uniform. Then, we propose a system for estimating workers' posture using a machine learning-based estimation algorithm and a data augmentation method corresponding to the sensing devices.

3 Proposed method

3.1 Overview of the proposed method

In this study, we propose a behavior estimation method for construction workers that uses a helmet-mounted terminal, which has already been installed on actual construction sites. Figure 1 shows an overview of the proposed system. The terminal with a built-in composite sensor provides sensing data including acceleration and positional information obtained by positioning system every second. These data are stored in the terminal's memory. The stored data are transferred to the database in the laptop by connecting the terminals to the equipment installed at the site. Then, we construct a system to estimate the postural state of construction workers from the stored sensing data by utilizing a neural network model that is trained using past behavior data of workers.

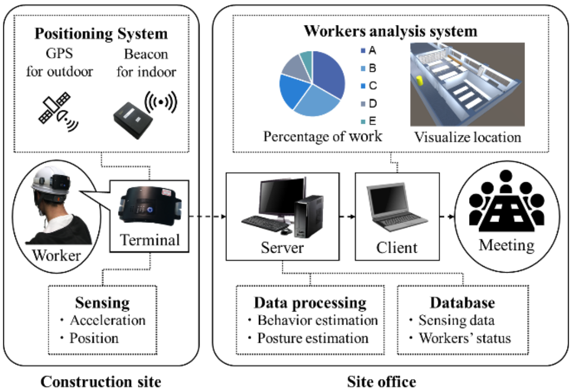

Figure 1. Overview of the proposed system

3.2 Data Acquisition by terminals

We use the "Construction Site Operation Monitoring System [16]" developed by Hitachi, Ltd. as a terminal for acquiring sensing data. The positional information is provided by Beacon or Global Positioning System (GPS) and sensing data, including triaxial acceleration, barometric pressure, and temperature are stored in the terminal's internal memory every second. The stored data are transferred to a server PC by connecting to the cradle installed on the site at the end of workday. The collected data are used to confirm the walking path of each worker verifying whether or not there are dangerous movements through the aforementioned system.

The monitoring system has functions such as detection of approaching the dangerous area and falling, and calculation of worker's posture based on the triaxial acceleration values obtained every second, but it does not have a function to estimate worker's behavior considering the change of worker's condition over time. Therefore, it is not possible to use the results obtained from the estimation of worker's behavior for productivity analysis or safety management measures.

3.3 Behavior estimation using sensing data

3.3.1 Data processing flow

Focusing on the fact that the workers' behavior estimation function is not implemented in the monitoring system, the proposed system employs a posture estimation method based on acceleration data and its time-series. Figure 2 shows the flow of the behavior estimation process in the proposed system. After the processing starts, input data for estimation are read from workers lists and sensing data registered in the database. In the preprocessing phase, acceleration data are extracted and formatted to suit the input of estimation model. Then the posture state of the target worker is estimated and processed as the posture data of the worker at that time by inputting the extracted data into the estimation model inside the system.

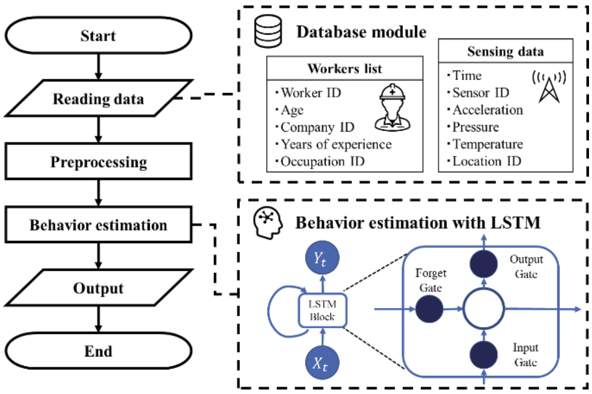

Figure 2. Data processing flow in the proposed system

3.3.2 Reading data and prerprocessing

The structure of the dataset used in this method for posture estimation and training of the estimation model is shown in equation (1). After reading sensing data from the database module, the time-series data is converted to the right side of the matrix in the equation (1). We assume that the worker's condition at time t is determined from the worker's condition before time t and the acceleration values around time t. Then, we define the dataset as a mapping between the posture label P_t at time t and the acceleration values around time t.

$$P_t \leftrightarrow \begin{bmatrix} x_{t-n} & y_{t-n} & z_{t-n} \\ \vdots & \vdots & \vdots \\ x_t & y_t & z_t \\ \vdots & \vdots & \vdots \\ x_{t+n} & y_{t+n} & z_{t+n} \end{bmatrix} \tag{1}$$

where

P_t: Posture label at time t

x_t, y_t, z_t: Acceleration value of each axis at time t

$2n + 1$: Time window width

3.3.3 Posture estimation with LSTM

In the posture estimation phase, a feature and classification algorithm should be selected in consideration of the terminals used for sensing and the behavior characteristics of individuals [4]. We implement a neural network model that automatically extracts and learns features from the input training data to perform posture estimation for various occupations and individual differences in this study. We also use Long Short-Term Memory (LSTM), which has been reported to have a high performance in predicting time series data among neural network models [17]. The structure of the LSTM employed in the system is shown in Figure 3. LSTM is a kind of recurrent neural network, which replaces the hidden layer in the recurrent neural network (RNN) with a module called LSTM blocks. LSTM addresses the gradient vanishing problem, where the value of the hidden layer is decayed and lost, which is inherited by the next time layer by adopting the LSTM block. It is known that LSTM performs better than RNN for the problem of predicting time series data with long-term dependency [18].

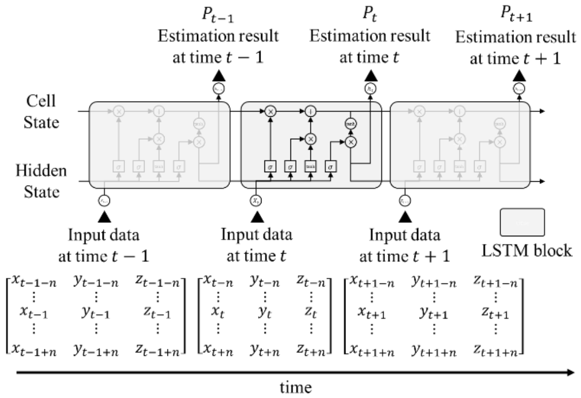

Figure 3. The structure of LSTM network for behavior estimation

3.3.4 Data augmentation

The data labeling should be done referring to data other than sensing data, such as video data, as shown in Figure 4, in creating datasets for behavior estimation. In the labeling of time-series behavior data, video data are commonly used to label behavior conditions [19]. On the other hand, this method is costly and labor-intensive, and inadvertent mislabeling may occur. In order to eliminate these tasks, previous works sharing human behavior data focusing on basic actions in daily life have been done [20]. Nevertheless, studies providing behavior data on specific workers such as construction workers have yet not been made. Therefore, we implement a data augmentation method for sensing data and propose a method to utilize limited data collected in the field as training data.

The conceptual diagram of the data augmentation proposed in this study is shown in Figure 5. While most of the studies on action recognition using inertial sensors use a sampling frequency of 25 Hz or higher, the terminals used in the proposed system have a low sampling frequency of 1 Hz. Taking this condition into consideration, the proposed method restores the acceleration waveform using the interpolation formula for the data discretized by sensing. After that, the number of data sets is artificially increased by cutting out data from the interpolated acceleration waveform at equal intervals. By using this method, it is possible to add diversity to the training data while maintaining the correspondence between time-series information and posture labels, which can be used to generate models with high generalization performance. Indeed, our low-frequency sampling system and this method are inappropriate for instantaneous motion classification problems such as gesture recognition. However, we adopt this method because of the postures targeted in this study record more stable data in the long term than gestures.

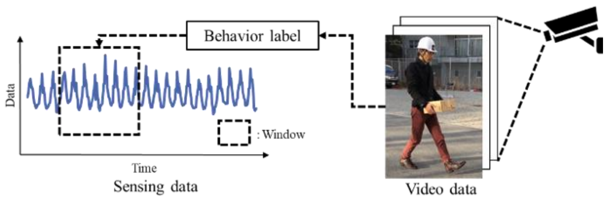

Figure 4. Conceptual diagram of data labeling

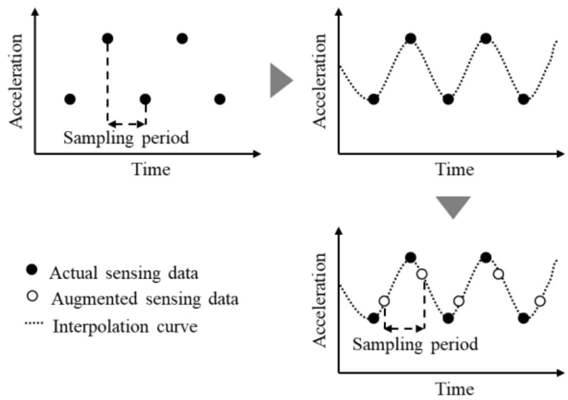

Figure 5. Conceptual diagram of data augmentation for sensing data

4 Verification of the proposed system

We evaluated the accuracy of posture estimation, verified the effectiveness of data augmentation, and estimated the number of binding points of rebar using acceleration data for a rebar worker at an actual construction site.

4.1 Experiment

4.1.1 Target work setting and data collection

A reinforcing bar worker (male, in his 50s, 37 years of experience) who works on binding rebars of floor slabs was asked to wear a terminal and recorded the working state together with the sensing by the terminal. In the work of binding the reinforcing bars of the floor slab, because the transition from the standing state to the forward-leaning posture and the squatting posture occurred repeatedly, the burden on the lumbar region of the worker was large, which caused a disorder of the musculoskeletal system. Because our system can be used for planning preventive measures against musculoskeletal disorders by quantifying and estimating the posture state. This criterion was also utilized for verification.

The rebar binding work is a typical repetitive work of moving to the rebar binding point, preparing a binding wire at a hand, transitioning to the forward-leaning or squatting posture, binding rebar, and moving to the next binding point. Figure 6 shows the flow of the work for binding the reinforcing bars of the floor slab. Unevenness and excessive reduction in repetitive work times are important information for ensuring an appropriate working environment for workers.

In this experiment, forward-leaning and squatting postures for a long time would be counted as rebar binding. The relationship between the posture and the rebar binding work of the floor slab is defined as shown in Figure 7. The number of rebar binding points is estimated based on that. Subsequently, by counting the number of binding points visually based on the actual video data and comparing the estimation results with the actual number, we consider the possibility of applying the proposed system to estimate the number of binding points of the reinforcing bars.

4.1.2 Datasets preparation

After acquiring the terminal record, the correct posture labels (hereinafter called ground truth label) were extracted from the video in every second. Then, the sensing data and correct labels were associated with each other to create data sets. Typically, the time window width is set to have a 50% overlap with respect to the sampling frequency [6], but in this experiment, given the low sampling frequency, we set it to 5 seconds with an 80% overlap. Then, the acceleration data for 5 seconds before and after was associated with the posture label P_t at time t. Three types of postures, standing, leaning-forward, and squatting, were extracted as the static state from "The Nagamachi Work Posture Classification [21]", which defines postures from the ergonomics point of view based on the magnitude of the load applied to the lumbar region of the human body in each work posture. For dynamic states, we set two categories: transitioning of posture states such as forward-leaning posture from standing posture and walking. Table 1 summarizes the aforementioned posture categories and their definitions in this experiment.

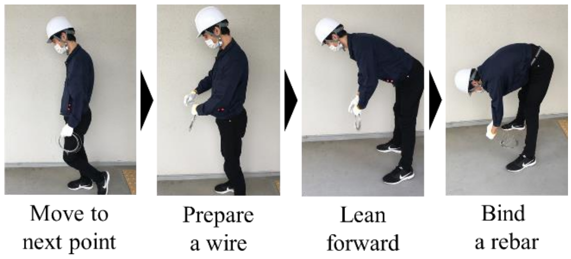

Figure 6. The workflow for binding the reinforcing bars of the floor slab

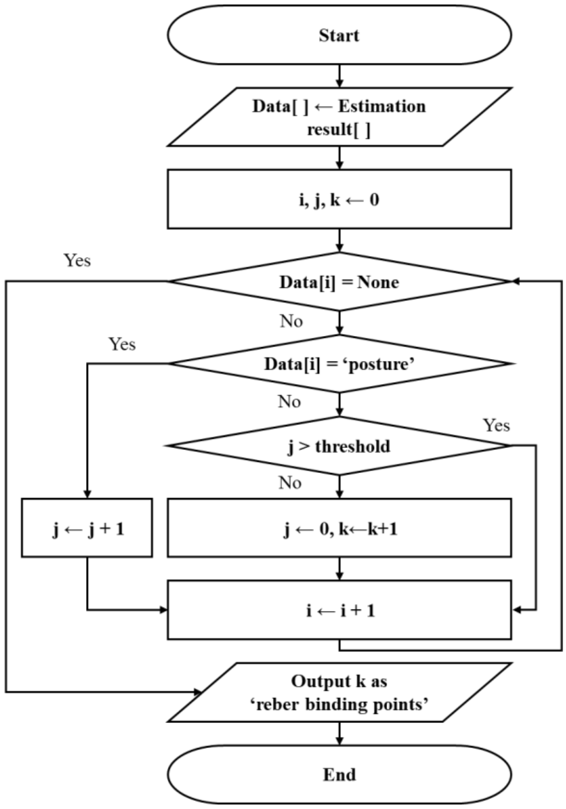

Figure 7. The flow of the estimation for the number of rebar binding points

Table 1. Posture classification of workers in the verification experiment

State	Posture	Definition
Static state	Standing	Upright posture with knees extended
	Forward-leaning	Posture with the waist bent by 30° or more with knees extended
	Squatting	Sitting on the heels posture
Dynamic state	Transitioning	State during changing posture
	Walking	State of advancing step by step

4.1.3 Estimation of accuracy indicators

After the data sets were created, they were divided into training data and test data. The estimation model was trained based on the training data, and the estimation accuracy was calculated based on the test data. Table 2 shows the breakdown of these data in the verification experiments. We allocated the correct labels and estimation results to the confusion matrix shown in Table 3 and calculated the precision, recall, F-measure, and accuracy from Equation (2) to Equation (5) to evaluate the accuracy of the posture estimation.

Table 2. The breakdown of datasets used for accuracy verification

Posture	Training data	Test data
Standing	160	16
Forward-leaning	607	40
Squatting	502	46
Transitioning	344	53
Walking	106	6
Total	**1719**	**161**

Table 3. Confusion matrix in two-class classification

		Estimation result	
		Positive	Negative
Ground truth	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

4.2 Results

4.2.1 Evaluation of data augmentation

Table 4 shows the estimation results with the data augmentation ratio as a comparison when generating the estimation model. For the data augmentation ratio, 1x (without data augmentation), 2x, 4x, and 8x values were used, while linear interpolation was also implemented with the waveforms for the augmentation. The results in Table 4 indicated that the F-measures in each posture, increased and decreased as the ratio of the data augmentation was increased, while the accuracy tended to increase as the ratio was increased. It was confirmed that forward-leaning and squatting postures could be estimated with the F-measure of 79% or higher for all data augmentation ratios.

Table 4. The estimation accuracy at each data augmentation ratio

	Posture	Data augmentation ratio			
		1x	**2x**	**4x**	**8x**
F-measure	Standing	0.73	0.60	0.62	0.60
	Forward-leaning	0.82	0.88	0.85	0.87
	Squatting	0.80	0.79	0.82	0.83
	Transitioning	0.41	0.51	0.60	0.65
	Walking	0.40	0.14	0.33	0.22
Accuracy		**0.68**	**0.70**	**0.73**	**0.73**

4.2.2 Evaluation in time-series

Figure 8 shows a time-series evaluation of the ground truth labels and the estimation results at the data augmentation ratio of 8x, which recorded the highest accuracy in the estimation results as shown in Table 4. It was confirmed that the accuracy increased when the same posture continued for a long time. On the other hand, the accuracy tended to be lower if the posture changed frequently in a short period.

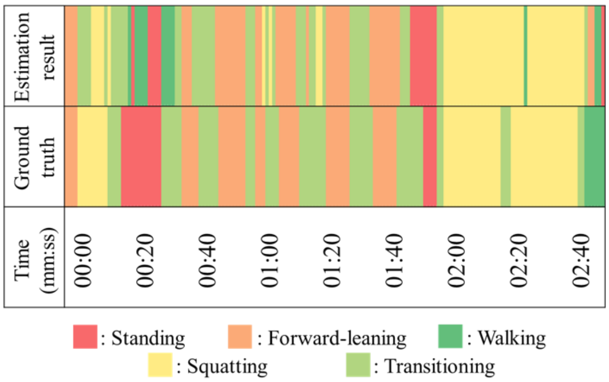

Figure 8. Time series evaluation of posture estimation

4.2.3 Evaluation of rebar binding points estimation

Table 5 shows the results of the estimation of the rebar binding points based on the results of the estimation flow of the binding points shown in Figure 7. The results of the posture estimation are shown in Figure 8. The condition that the rebar is bound while the forward-leaning or squatting posture was continued for more than 5 seconds was set. The number of binding points, the total time required for rebar binding, and the time spent on binding each point were also calculated.

Table 5. Evaluation of rebar binding points estimation

Input data for estimation	Posture estimation	Ground truth posture
The number of binding points	6	8
Total binding time (sec)	72.0	83.0
Binding time per point (sec)	12.0	10.4

4.3 Discussions

The results in Table 4 show that the proposed system can estimate the worker's posture correctly up to 73% accuracy based on the triaxial acceleration. The accuracy of the proposed system was increased from 68% to 73% by using the data augmentation method. It was also confirmed that the F-measure of 79% or more was able to be achieved for the forward-leaning and squatting postures. Their accuracies are higher than the other gestures in every data augmentation ratio. This is because the LSTM network learned the change in direction of the gravitational acceleration caused by the tilt of the worker's head as effective feature for posture estimation. In the posture classification based on the influence on the musculoskeletal system [21], forward-leaning and squatting postures are set as the level of 5th and 6th out of 10th respectively. Therefore, calculating the cumulative frequency of these levels enable improvement of working environment using quantitative indicators. For further improvement of accuracy, we need to use positional information that was not used in this experiment and add small devices that consider intrusiveness of wearing. According to the results shown in Table 5, the number of rebar binding points were estimated, and six of the eight points were correctly estimated from the posture estimation results. By comparing these estimated results with the quantities of components extracted from product models such as Building Information Modeling (BIM), site stakeholders will be able to understand the progress quantitatively.

5 Conclusions

This paper proposed a system for estimating workers' posture using a helmet-mounted terminal, which is already in use at a construction site to collect worker's behavior data. From the results of posture estimation using triaxial acceleration data acquired at the terminal, it was confirmed that five different postures could be estimated with an accuracy of up to 73% by using LSTM and the data augmentation method. In particular, the system was able to detect forward-leaning and squatting postures with high accuracy, which indicates the system can be used to improve the ergonomic work environment, such as quantifying the load on the body, by calculating the cumulative time of those postures. It was also confirmed that the results of the posture estimation can be used to predict the number of rebar binding points.

Future work includes collecting data to expand the range of jobs and postures to be estimated and improving the accuracy of posture estimation by linking with other types of sensing data such as positional information. We also aim to develop a management system that links behavior estimation with geometric and attribute information of the BIM model.

References

[1] Labour Standards Bureau, Ministry of Health, Labour and Welfare: Survey on Industrial Accidents, On-line: https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei11/rousai-hassei/dl/b19-16.pdf, Accessed: 6/6/2020.

[2] Ministry of Land, Infrastructure, Transport and Tourism: White Paper on Land, Infrastructure, Transport and Tourism in Japan 2019, On-line: https://www.mlit.go.jp/common/001325161.pdf, Accessed: 6/6/2020.

[3] Park, C. and Kim, H.: A framework for construction safety management and visualization system,

Automation in Construction, Vol.33, pp.95-103, 2013.

[4] Yang, X., Wang, F., Li, H., Yu, Y., Luo, X. and Zhai, X.: A Low-Cost and Smart IMU Tool for Tracking Construction Activities, In 2019 *Proceedings of the 36th ISARC*, Banff, Alberta, Canada, pp.35-41.

[5] Bao, L. and Intille, S.S.: Activity Recognition from User-Annotated Acceleration Data, *Pervasive Computing*, pp.1-17, 2004.

[6] Akhavian, R. and Behzadan, A.H.: Smartphone-based construction workers' activity recognition and classification, *Automation in Construction*, Vol.71, pp.198-209, 2016.

[7] Zhang, M., Chen, S., Zhao, X. and Yang, Z.: Research on Construction Workers' Activity Recognition Based on Smartphone, *Sensors*, Vol.18, 2667, 2018.

[8] Yan, X., Li, H., Li, A. and Zhang, H.: Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention, *Automation in Construction*, Vol.74, pp.2-11, 2017.

[9] Chen, L., Hoey, J., Nuget, C. D., Cook, D.J. and Yu, Z.: Sensor-Based Activity Recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol.42, No.6, pp.790-808, Nov. 2012.

[10] Nakamura, Y. and Kanade, T.: Semantic Analysis for Video Contents Extraction – Spotting by Association in News Video, In *Proceedings of the fifth ACM international conference on Multimedia*, pp.393-401, 1997.

[11] Yang, J., Arif, O., Vela, P.A., Teizer, J. and Shi, Z.: Tracking multiple workers on construction sites using video cameras, *Advanced Engineering Informatics*, Vol.24, No.4, pp.428-434, 2010.

[12] Luo, H., Xiong, C., Fang, W., Love, P.E.D., Zhang, B. and Ouyang, X.: Convolutional neural networks: Computer vision-based workforce activity assessment in construction, *Automation in Construction*, Vol.94, pp.282-289, 2018.

[13] Fang, W., Ding, L., Luo, H. and Love, P.E.D.: Falls from heights: A computer vision-based approach for safety harness detection, *Automation in Construction*, Vol.91, pp.53-61, 2018.

[14] Park, M. and Brilakis, I.: Construction worker detection in video frames for initializing vision trackers, *Automation in Construction*, Vol.28, pp.15-25, 2012.

[15] Ray, S.J. and Teizer, J.: Real-time construction worker posture analysis for ergonomics training, *Advanced Engineering Informatics*, Vol.26, No.2, pp.439-455, 2012.

[16] Hitachi, Ltd.: Hitachi Develops Systems to Increase Productivity and Safety at Construction Sites Using IoT Technologies, On-line: http://www.hitachi.com/New/cnews/month/2018/03/180327.pdf, Accessed: 6/6/2020.

[17] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Computation*, Vol.9, No.8, pp.1735-1780, 1997.

[18] Graves, A., Fernández, S. and Schmidhuber, J.: Bidirectional LSTM networks for improved phoneme classification and recognition, In *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II*, pp.799–804, Sep. 2005.

[19] Anguita, D., Ghio, A., Cneto, L., Parra, X. and Reyes-Ortiz, J.L.: A Public Domain Dataset for Human Activity Recognition Using Smartphones, ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp.437-442.

[20] Ramanan, D. and Forsyth, D.A.: Automatic Annotation of Everyday Movements, In *Proceedings of Neural Info. Proc. Systems (NIPS)*, Vancouver, Canada, Dec. 2003.

[21] Nagamachi, A.: Requisite and practices of participatory ergonomics, *International Journal of Industrial Ergonomics*, Vol.15, pp.371-377, 1995.