Automated Productivity Estimation of Masonry Work using a Deep Learning Approach and Wearable Motion Sensors

Kinam Kim and Yong K. Cho, Ph.D.

Abstract— **The construction industry has faced a lagging improvement or decline, while other sectors have shown a noticeable improvement in productivity. Despite the importance of managing productivity, current practices rely heavily on manual observation. Such approaches are labor-intensive and error-prone. To facilitate the automated productivity monitoring of individual workers, this study proposes a framework for estimating the labor productivity of workers using wearable motion sensors and a deep learning approach. The framework consists of two modules, including an activity recognition module and a productivity estimation module. In the activity recognition module, a long short-term memory (LSTM) network is implemented to recognize the activities of workers performing a bricklaying task. In the productivity estimation module, the recognized activities are discretized using a dynamic time warping (DTW) so that the activities can be counted. Based on the predefined workflow, the production amount is estimated, and finally, the labor productivity of individual workers is estimated.**

I. Introduction

Understanding the ever-changing behavior of construction workers is essential to achieving success in construction projects because construction projects are labor-intensive and rely heavily on manual tasks. This arouses a strong need to understand workers' activities to improve and ensure the safety and productivity of individual workers. To be specific, the construction industry has faced a lagging improvement or even a decline in productivity, while other industries have shown a noticeable improvement in productivity [1]. Even though several sections, including multi-family housing and industrial construction, showed a notable labor productivity growth, these sections accounted for less than 10% of the total construction hours in 2012 [2]. Despite the importance of managing productivity in construction, current approaches to measure and manage productivity rely heavily on manual observation, which is labor-intensive and error-prone. Moreover, it is challenging to measure the productivity of multiple workers at the individual level using those approaches.

To address these challenges and facilitate individual-level monitoring of labor productivity, this study proposes a framework for estimating the labor productivity of individual workers without excessive observation. First, an activity recognition model is implemented using an LSTM network, which is one of the recurrent neural networks. This model classifies a sequence of motion sensor data into a particular activity among multiple activity classes of the bricklaying task. Next, the recognized activities are discretized using a DTW technique. Because the classification results of the LSTM network include duplicated activities depending on data preprocessing parameters, the DTW is adopted to exclude the duplicated activities by calculating the similarity between two time-series data so that the production amount, i.e., the number of bricks placed, can be estimated from the classification results. Finally, the labor productivity of individual workers is estimated based on the production amount. With the proposed framework, it is expected that the behavior and labor productivity of workers can be monitored without excessive manual observation, and finally, such information can be used for managerial applications in construction projects to improve productivity and optimize the relevant resources.

II. Related Work

This section reviews motion sensor-based motion and activity recognition methods. In general, state-of-the-art motion and activity recognition methods use machine learning algorithms to recognize motion and activity patterns from the sensor data.

Motion and activity recognition methods were utilized to identify various motions and activities using machine learning algorithms [3]–[9]. Four activities of construction workers were recognized using five types of machine learning algorithms [10]. This study utilized a smartphone attached to the arm to collect motion sensor data. Likewise, a wristband-type accelerometer sensor was used to recognize the activities of a masonry worker using machine learning algorithms [6], [7].

Regarding safety, motion and activity recognition can be utilized to detect workers' unsafe postures. A support vector machine (SVM) classifier with a supervised motion tensor decomposition was developed to identify the awkward postures of workers [11]. Similarly, near-miss falls were detected using one-class SVM with motion sensor data from wearable inertial measurement units (IMUs) [12]. Gait stability can be analyzed using IMUs attached to the ankles for detecting fall risks [13]–[15]. A deep-learning approach, i.e., convolutional long short-term memory (CNN-LSTM), was presented to recognize the motions of workers using five IMUs [5]. The presented method recognized the motions that can cause musculoskeletal disorders, including bending and kneeling.

Motion and activity recognition methods were utilized to calculate the productive time duration of workers and analyze their productivity [16], [17]. In these studies, productive time

K. Kim is with the Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: kkim734@gatech.edu).

Y. K. Cho, Ph.D. is with the Georgia Institute of Technology, Atlanta, GA 30332 USA (Corresponding author, phone: 402-385-2038, e-mail: yong.cho@ce.gatech.edu).

was calculated by counting the number of samples classified by a machine learning classifier using a smartphone with an embedded IMU. Masonry workers were grouped into the expert and inexpert groups using a support vector machine classifier based on their motions for comparing their productivity [18].

Although there have been several research efforts that show great potential for recognizing the motions and activities of workers, a few efforts have been made to automatically estimate the labor productivity of individual workers without manual observation. Measuring productivity is heavily relied on manual observation based on sampling work and idle times. However, this approach is labor-intensive, error-prone, and unable to provide individual workers' productivity. Even if productivity can be indirectly obtained by measuring work and idle times, it does not essentially represent the number of work placed by each worker.

III. Methodology

A framework that estimates the labor productivity of individual workers using an activity discretization technique is developed by 1) recognizing the activities of individual workers, 2) estimating the production amount by discretizing and counting the activities, and 3) estimating their labor productivity.

A. Activity Recognition Model Implementation

A bricklaying task is selected as the target task for developing the framework as this task involves repetitive activities, and a unit of productivity is directly measurable, i.e., the number or area of bricks placed per unit hour. Figure 1 presents a process of the bricklaying task at the level of decomposed activities. The bricklaying task includes brick flow and mortar flow [19]. Brick flow is a process of handling bricks, leveling tools, and tooling brick joints. Mortar flow is a process of creating mortar, delivering mortar, checking the quality of mortar, and pouring mortar. In this study, the scope of predicting labor productivity is limited to the brick flow, including lifting, carrying, placing a brick, and tooling. While these activities are considered variable activities, other activities are considered fixed activities.

Activity datasets are collected from subjects performing a bricklaying task. Target activity classes include walking without a brick, lifting, carrying, placing a brick, tooling, and idling. As the task mainly involves hand-oriented motions and activities, two IMUs located on the hip and hand, depending on their handedness, are utilized to collect the sensor data because two IMUs at a certain distance can represent the whole body motions [3].

Each IMU generates a feature set composed of 9 values: acceleration (3 values for x, y, and z axes), angular velocity (3 values for x, y, and z axes), and magnetic field (3 values for x, y, z axes), and two feature sets from two IMUs are concatenated to form an input vector; thus one input vector includes 18 values.

Once the datasets are generated, an activity recognition algorithm is developed by using an LSTM network. The two-stacked LSTM network developed in [3] is adopted to implement the activity recognition model. While the same structure is adopted, the scope of recognition is set to the classification of the activities related to the bricklaying task as shown Fig. 2. Hence, the input vectors are labeled at the activity level. Since an LSTM network is capable of learning sequential information from data, activities can also be effectively recognized by using the LSTM network.

B. Labor Productivity Estimation

The activity recognition allows adjacent activities to be overlapped with each other when activity sequences are generated. Although this technique is important when AI-based classification is developed to minimize the noisy signal and enhance the distinguishability of the data [6], it is not able to identify how much relevant work is made directly. Existing approaches have tried to calculate productivity indirectly by estimating the time duration spent on specific activities. To address these challenges, this study adopts a Dynamic Time Warping (DTW) technique to discretize the recognized activities into separated activities so that the production amount can be directly estimated. DTW is a technique that finds the similarity between two time-series sequences by warping the time axis to align two sequences [20]. In DTW, the sequences are stretched or shrunk along the time axis non-linearly to identify the best match between two sequences. The more two sequences have the same patterns, the smaller DTW distance is measured. The DTW distance between two sequences is determined by using Equation (1).

$$D(i,j) = \min\{D(i-1,j-1), D(i-1,j), D(i,j-1)\} + |x_i, y_j| \quad (1)$$

where $D(i,j)$ is the DTW distance between two sequences $x[1:i]$ and $y[1:j]$ [21]. This equation iteratively calculates the distance between elements in two sequences until all elements are compared.

To discretize the recognized activities where overlapping is allowed, DTW distances within activities in the same activity class are calculated. After input vector sequences of the recognized activities are linearized, the DTW distances are calculated between adjacent sequences if they are classified as the same activity. This calculation is repeated until a new class is observed. If the calculated DTW distance is smaller than a threshold value, the later sequence is excluded from counting

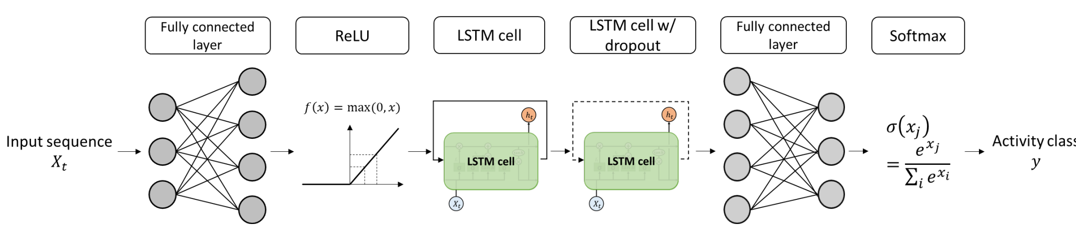

Figure 1. Two-stacked LSTM network structure [3].

activities. By repeating this procedure, the recognized activities are discretized so that the separated activities can be counted.

In this study, the unit of productivity of the bricklaying is defined as the number of bricks placed per unit hour. The production amount of each worker is then estimated by detecting a lifting activity followed by a placing activity. The particular sequential window is considered to detect a placing activity that follows a lifting activity because there might be other activities between them, such as carrying and idling. With these definitions and the estimated production amount, the productivity is calculated using Equation (2).

$$Productivity = \sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{m} Work\ placed_j}{Work\ hours} \right)_i \quad (2)$$

where n is the number of workers in the same task and j is the number of the estimated work placed by each worker.

C. Implementation and Result

The experiments were conducted in an outdoor environment similar to an actual jobsite. Fig. 2 presents examples of target activities in the experiments. The total number of bricks to be laid was 60, and 30 bricks were delivered to the bricklaying area at a time. To collect the IMU data from the subjects, data collecting devices, as shown in Fig. 3, developed by the Robotics and Intelligent Construction Automation Laboratory (RICAL) group at the Georgia Institute of Technology, were utilized. The devices were carried by workers wearing a watch-type device depending on their handedness and safety vests with pockets on their lower back. The devices are equipped with a wireless communication module for Wi-Fi and Bluetooth, a micro processing unit, data storage, battery, and an IMU. The IMU used in this study consisted of three triaxial sensors, including an accelerometer, gyroscope, and magnetometer, which have digital resolutions of 0.98 mg, 0.004°/s, and 0.3 μT, respectively.

Two datasets were collected from the subjects who performed bricklaying. Subjects performing bricklaying had different behavioral patterns because the brick piles were placed in different locations, which led to different routines for laying bricks. Thus, datasets were separately collected from the subjects. The datasets contain 7277 and 8437 data points, respectively. The collected data points were labeled with timestamps to be manually compared with the recorded videos later.

Figure 2. Examples of target activities; (a) walking without a brick, (b) lifting a brick, (c) carrying a brick, (d) placing a brick, (e) tooling, and (f) being idle.

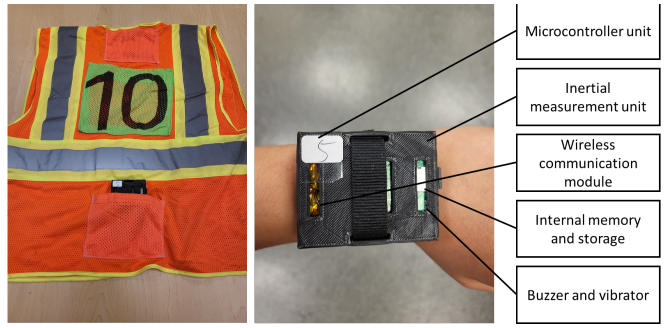

Figure 3. Data collecting device and a safety vest with pockets.

As a result, the six networks showed accuracies of 88.00% and 90.94% on the testing data, respectively. With the recognized activities, DTW distances were calculated to discretize the activities. The threshold values of DTW distances were determined for each classification result. Once the DTW distance between two adjacent sequences of activities is calculated, the later sequence is excluded if the distance is smaller than the threshold value. This indicates two sequences represent a single activity and should be counted as one repetition. This pairwise comparison is conducted throughout the whole dataset. As a result, one-dimensional vectors containing the discretized activities were derived. In these vectors, the production amount can be estimated by detecting lifting activities followed by placing activities. Particular sequential windows were allowed between lifting and placing activities because other activities, such as carrying, unexcluded lifting, or misclassified activities, can exist between the activities of interest. With the estimated production amount, labor productivity is estimated as shown in Table 1. In this study, mortar flow was not considered as variable time but fixed time. Spreading mortar should be done between each bricklaying. Thus, the average cycle time of spreading mortar, which is 2.5 seconds [6], was added to the productivity prediction as a constant value. This was done by adjusting the trial duration depending on the estimated production amount, i.e., the number of bricks placed. As a result, 18.18% and 3.1% of productivity estimation errors were derived.

Table 1 Labor productivity of individual workers.

Subject	Role	Original Duration (minutes)	Estimated Production Amount (bricks or barrows)	Adjusted Duration (minutes)	Productivity (bricks/hour or barrows/hour)	Ground Truth (Production Amount)	Ground Truth (Productivity)
1	Laying	6.65	27	7.78	208.23	33	254.50
2	Laying	6.65	26	7.73	201.81	27	208.23

IV. Conclusion

This study proposed a framework for estimating the labor productivity of individual workers by using a deep learning-based activity recognition model and an activity discretization technique. The experiments showed that 3.1% to 18.18% of estimation errors were achieved by the proposed method. It is expected that the labor productivity of individual workers can be estimated and monitored without excessive manual observation with the framework.

Future studies will focus on conducting case studies to validate the practical and technical feasibility of the framework. Moreover, further research will be conducted to develop decision support methods that allocate workforces and optimize cost and duration based on their productivity.

V. ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. (#1919068). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

REFERENCES

[1] McKinsey&Company, "The construction productivity imperative," *McKinsey&Company*, 2015. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/the-construction-productivity-imperative (accessed Jan. 22, 2018).

[2] L. Sveikauskas, S. Rowe, J. Mildenberger, J. Price, and A. Young, "Measuring productivity growth in construction," *Mon. Labor Rev.*, Jan. 2018, doi: 10.21916/mlr.2018.1.

[3] K. Kim and Y. K. Cho, "Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition," *Autom. Constr.*, vol. 113, p. 103126, May 2020, doi: 10.1016/j.autcon.2020.103126.

[4] K. Kim and Y. K. Cho, "Automatic Recognition of Workers' Motions in Highway Construction by Using Motion Sensors and Long Short-Term Memory Networks," *J. Constr. Eng. Manag.*, vol. 147, no. 3, p. 04020184, Dec. 2020, doi: 10.1061/(ASCE)CO.1943-7862.0002001.

[5] J. Zhao and E. Obonyo, "Convolutional Long Short-Term Memory Model for Recognizing Postures from Wearable Sensor," 2019, Accessed: Jan. 15, 2020. [Online]. Available: http://ceur-ws.org/Vol-2394/paper36.pdf.

[6] J. Ryu, J. Seo, H. Jebelli, and S. Lee, "Automated Action Recognition Using an Accelerometer-Embedded Wristband-Type Activity Tracker," *J. Constr. Eng. Manag.*, vol. 145, no. 1, pp. 1–14, 2019, doi: 10.1061/(ASCE)CO.1943-7862.0001579.

[7] J. Ryu, J. Seo, M. Liu, S. Lee, and C. T. Haas, "Action Recognition Using a Wristband-Type Activity Tracker: Case Study of Masonry Work," in *Construction Research Congress 2016*, 2016, pp. 790–799, Accessed: Apr. 01, 2019. [Online]. Available: https://ascelibrary.org/doi/pdf/10.1061/9780784479827.080.

[8] X. Su, H. Tong, and P. Ji, "Activity recognition wit h smartphone sensors," *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 235–249, Jun. 2014, doi: 10.1109/TST.2014.6838194.

[9] E. Valero, A. Sivanathan, F. Bosché, and M. Abdel-Wahab, "Analysis of construction trade worker body motions using a wearable and wireless motion sensor network," *Autom. Constr.*, vol. 83, pp. 48–55, Nov. 2017, doi: 10.1016/J.AUTCON.2017.08.001.

[10] R. Akhavian and A. H. Behzadan, "Smartphone-based construction workers' activity recognition and classification," *Autom. Constr.*, vol. 71, pp. 198–209, Nov. 2016, doi: 10.1016/J.AUTCON.2016.08.015.

[11] J. Chen, J. Qiu, and C. Ahn, "Construction worker's awkward posture recognition through supervised motion tensor decomposition," *Autom. Constr.*, vol. 77, pp. 67–81, May 2017, doi: 10.1016/J.AUTCON.2017.01.020.

[12] K. Yang, C. R. Ahn, M. C. Vuran, and S. S. Aria, "Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit," *Autom. Constr.*, vol. 68, pp. 194–202, Aug. 2016, doi: 10.1016/j.autcon.2016.04.007.

[13] H. Jebelli, C. R. Ahn, and T. L. Stentz, "Comprehensive Fall-Risk Assessment of Construction Workers Using Inertial Measurement Units: Validation of the Gait-Stability Metric to Assess the Fall Risk of Iron Workers," *J. Comput. Civ. Eng.*, vol. 30, no. 3, 2015, doi: 10.1061/(ASCE)CP.1943-5487.0000511.

[14] K. Yang, C. R. Ahn, and H. Kim, "Validating ambulatory gait assessment technique for hazard sensing in construction environments," *Autom. Constr.*, vol. 98, pp. 302–309, Feb. 2019, doi: 10.1016/j.autcon.2018.09.017.

[15] K. Yang and C. R. Ahn, "Inferring workplace safety hazards from the spatial patterns of workers' wearable data," *Adv. Eng. Informatics*, vol. 41, p. 100924, Aug. 2019, doi: 10.1016/j.aei.2019.100924.

[16] R. Akhavian and A. H. Behzadan, "Productivity Analysis of Construction Worker Activities Using Smartphone Sensors," in *Proceedings of 16th International Conference on Computing in Civil and Building Engineering*, 2016, pp. 1067–1074, Accessed: May 07, 2018. [Online]. Available: https://pdfs.semanticscholar.org/9cc3/2a15b75765e977bc1b8ccc6d545e8c5168a2.pdf.

[17] N. D. Nath and A. H. Behzadan, "Construction Productivity and Ergonomic Assessment Using Mobile Sensors and Machine Learning," in *Computing in Civil Engineering 2017*, Jun. 2017, pp. 434–441, doi: 10.1061/9780784480847.054.

[18] A. Alwasel, A. Sabet, M. Nahangi, C. T. Haas, and E. Abdel-Rahman, "Identifying poses of safe and productive masons using machine learning," *Autom. Constr.*, vol. 84, pp. 345–355, Dec. 2017, doi: 10.1016/j.autcon.2017.09.022.

[19] S. Alireza Abbasian-Hosseini, A. Nikakhtar, and P. Ghoddousi, "Verification of Lean Construction Benefits through Simulation Modeling: A Case Study of Bricklaying Process," *KSCE J. Civ. Eng.*, vol. 18, no. 5, pp. 1248–1260, 2014, doi: 10.1007/s12205-014-0305-9.

[20] E. J. Keogh and M. J. Pazzani, "Derivative Dynamic Time Warping," in *Proceedings of the 2001 SIAM international conference on data mining*, Apr. 2001, pp. 1–11, doi: 10.1137/1.9781611972719.1.

[21] J. Trelinski and B. Kwolek, "CNN-based and DTW features for human activity recognition on depth maps," *Neural Comput. Appl.*, vol. 33, pp. 14551–14563, 2021, doi: 10.1007/s00521-021-06097-1.