Evaluation of Machine Learning Algorithms for Worker's Motion Recognition using Motion Sensors

Kinam Kim,[1] Jingdao Chen,[2] and Yong K. Cho, Ph.D., M.ASCE[3]

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA, 30332-0355; e-mail: kkim734@gatech.edu
[2]Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA, 30332-0355; e-mail: jchen490@gatech.edu
[3]School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA, 30332-0355; e-mail: yong.cho@ce.gatech.edu

ABSTRACT

Construction tasks involve various activities composed of one or more body motions. It is essential to understand the dynamically changing behavior and state of construction workers to manage construction workers effectively with regards to their safety and productivity. While several research efforts have shown promising results in activity recognition, further research is still necessary to identify the best locations of motion sensors on a worker's body by analyzing the recognition results for improving the performance and reducing the implementation cost. This study proposes a simulation-based evaluation of multiple motion sensors attached to workers performing typical construction tasks. A set of 17 Inertial Measurement Unit (IMU) sensors is utilized to collect motion sensor data from an entire body. Multiple machine learning algorithms are utilized to classify the motions of the workers by simulating several scenarios with different combinations and features of the sensors. Through the simulations, each IMU sensor placed in different locations of a body is tested to evaluate its recognition accuracy toward the worker's different activity types. Then, the effectiveness of sensor locations is measured regarding activity recognition performance to determine relative advantage of each location. Based on the results, the required number of sensors can be reduced maintaining the recognition performance. The findings of this study can contribute to the practical implementation of activity recognition using simple motion sensors to enhance the safety and productivity of individual workers.

INTRODUCTION

Monitoring the behavior and working state of construction workers is a challenging task due to the dynamic nature of construction projects. Since construction tasks involve various physical activities consisted of one or more body motions, understanding the ever-changing activities and motions of the workers is necessary to manage the workers effectively in order to improve safety and productivity. Construction projects, in general, require excessive and repetitive physical activities, which arouses the strong need for understanding the worker's activities and motions for ensuring and improving the safety, health, and productivity of individual workers.

To recognize motions of the construction workers, machine learning algorithms have been utilized to classify motion sensor data. Several research efforts have shown promising action recognition performance using the motion sensor data. However, there is a lack of understanding about appropriate locations of motion sensors on a worker's body to better recognize the motions and activity classifications. Motion sensor data, such as acceleration and angular velocity, varies widely depending on its location on the body. Thus, the understanding the locations of each sensor and its impact on motion recognition can be used to improve the recognition performance and reduce the implementation cost. Hence, this study proposes a simulation-based evaluation of motion sensor locations on a construction worker's body. Seventeen inertial measurement unit (IMU) sensors are used to capture the motions from the entire body while the worker is performing typical construction tasks. Five machine learning algorithms are utilized to classify the motions. By simulating several scenarios with different combinations and features of the sensors, each IMU sensor located on different parts of the body is evaluated, and the locations with higher recognition rates are determined.

LITERATURE REVIEW

Two approaches have been widely utilized to recognize human's motions and activities, which are 1) image-based approach and 2) sensor-based approach. While the image-based approach collects motion data by extracting feature points or a 2D or 3D skeleton model from images, the sensor-based approach collects motion data from sensors including accelerometer, gyroscope, and magnetometer which are usually located on body joints.

Several research studies utilize RGB-D cameras, e.g., Kinect, which provide depth information that can be used to build 2D or 3D skeleton model of a human (Escorcia et al. 2012; Han and Lee 2013; Michel et al. 2017; Yu et al. 2017). A 2D skeleton model developed from Kinect images was used to recognize the leading postures of unsafe behaviors (Yu et al. 2017). Assuming that leading postures play a role as the precursor of unsafe behaviors, the ranges of joint angles of the leading postures were determined through an experimental study. Multiple cameras, 3D camcorder, and Kinect were used to generate the 3D skeleton model (Han and Lee 2013). By comparing the skeleton model and motion templates, unsafe behaviors during ladder climbing were recognized. A machine learning algorithm, e.g., the support vector machine classifier, was deployed to recognize construction workers' actions using a 3D skeleton model generated by Kinect (Escorcia et al. 2012). While RGB-D camera or multiple cameras provide useful information for motion recognition, several efforts have been made to achieve the same purpose using a single camera (Kim et al. 2016; Peddi et al. 2009; Yang et al. 2016). Machine learning classifiers using image-based descriptors were developed to recognize various construction activities (Yang et al. 2016). Although this study did not show satisfactory performance for every activity, the study provided an insight into how images from a single camera can be used for construction activity recognition. However, the image-based approach still has a limitation that occlusion between objects can result in incomplete object detection which may lead decrease of a recognition rate.

For sensor-based approaches, smartphones are widely used because smartphones have embedded sensors for recognizing motions including accelerometer and gyroscope, and it is a cost-effective way to collect data (Akhavian and Behzadan 2016; Bayat et al. 2014; Kwapisz et al. 2011; Nath et al. 2017; Yang 2009). The study by Akhavian and Behzadan (2016) analyzed the productivity of construction worker through a machine learning classification method using data from accelerometer and gyroscope embedded in a smartphone. The study measures the time spent on each activity using the classifier and analyzes productivity. Other studies solely used acceleration data to train machine learning classifiers for recognizing human activities (Bayat et al. 2014; Kwapisz et al. 2011; Yang 2009). Although these studies showed promising classification performance, target activities are limited to daily activities such as running, dancing, and walking which are not descriptive enough to fully recognize specific construction tasks. A study by Nath et al. (2017) used two smartphones and embedded accelerometers and gyroscopes to analyze the construction worker's ergonomic posture while the worker is performing screw driving tasks. On the other hand, Valero et al. (2017) used a separately developed sensor system to identify angular thresholds for detecting motions of construction workers. The study utilizes multiple sensors so that the data can represent motions in more detail. Cho et al. (2018) utilized motion sensors to detect the unsafe posture of exoskeleton wearers. However, there is still a lack of understandings of how to determine locations and the number of sensors to recognize the construction worker's motions.

METHODOLOGY

The proposed study follows the procedure shown in Figure 1. Each step of the procedure will be explained in the following subsections.

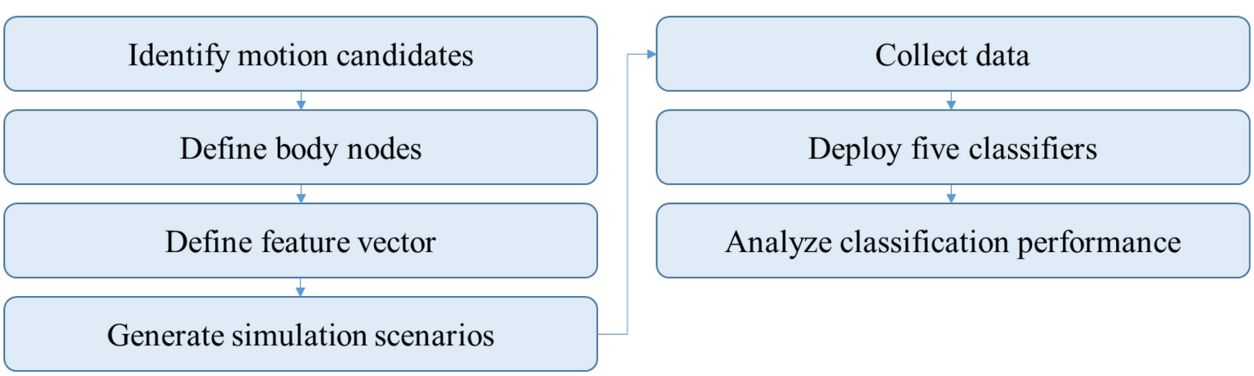

Figure 1. Evaluation procedure of sensor locations.

Identify motion candidates

The following motions for typical construction tasks are selected as motion candidates; standing, bending, squatting, walking, twisting, working overhead, kneeling, and using stairs. Each of the bending, squatting, and kneeling activities is divided into three classes (such as bending-up, bending-down, and bending) to reduce the loss of information caused by transitions of motions. For example, bending-up and bending-down motions indicate transitional motions from the bending motion to other motions or vice versa. Hence, fourteen motion classes are used in this study.

Define body nodes

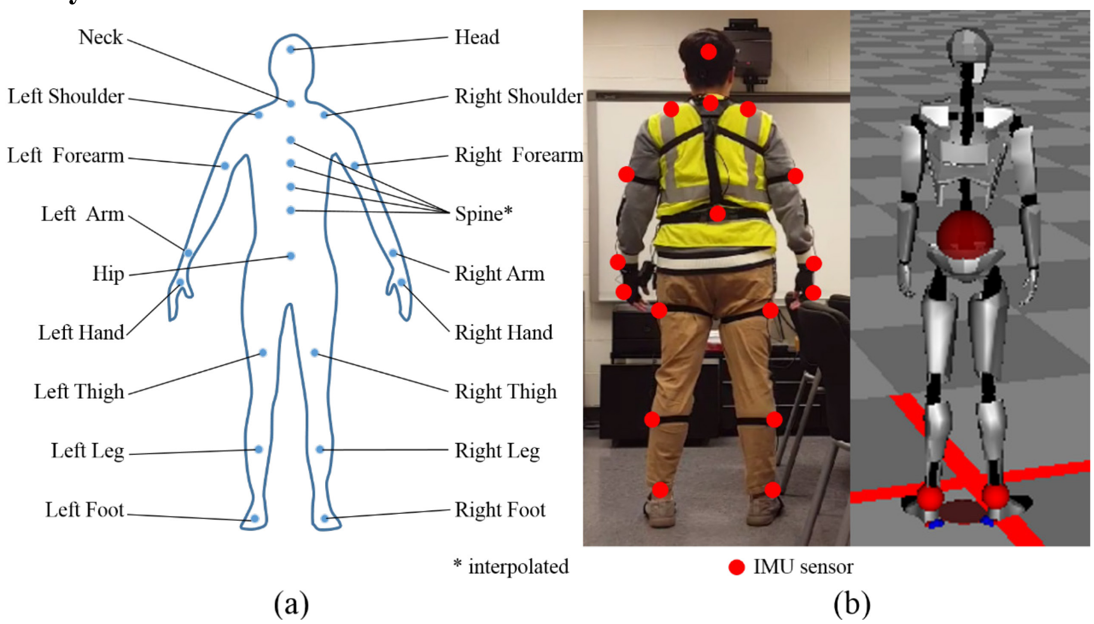

Figure 2. (a) Body nodes and (b) wearable sensor locations.

As shown in Figure 2 (a), 21 body joints or body parts are designated as nodes for simulation. Among the 21 nodes, 4 nodes located on the spine can be tracked by interpolation between the neck and hip nodes. Thus, the integrated wearable sensor system only requires 17 IMU sensors (red dots in Figure 2 (b)) to collect data from all the nodes.

Define feature vector

Once the nodes are defined, feature vectors are constructed using data from three types of sensors (accelerometer, gyroscope, and magnetometer), embedded in each IMU. Each feature vector is used as input of machine learning classifiers. As shown in Figure 3, each node with a single IMU generates a feature set composed of 13 values including quaternion (4 values), acceleration (3 values), velocity (3 values), and angular velocity (3 values). Feature sets from multiple nodes are concatenated to construct an overall feature vector. The generated feature vectors are used as training and test data for machine learning classifiers. Each data contains a discrete and time-independent motion state.

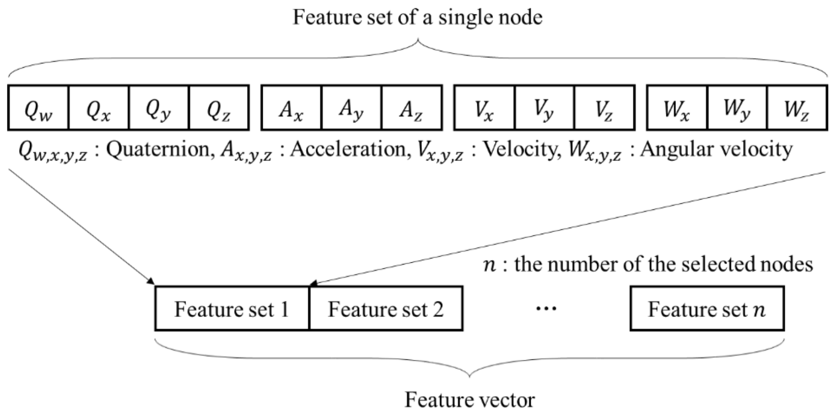

Figure 3. Formation of the feature vector.

Generate simulation scenarios

To identify the impact of the sensor locations on motion recognition performance, simulation scenarios are generated by considering different combinations of the selected nodes. Table 1 shows the combinations of nodes used in the simulation. Each scenario uses a different dimension for the feature vector depending on the number of selected nodes.

Table 1. Combinations of nodes used in the simulation

Combination	Selected nodes (The number of nodes)	Combination	Selected nodes (The number of nodes)
1	All nodes (21)	17	Right Foot (1)
2	Upper body (15)	18	Left Thigh (1)
3	Lower body (7)	19	Left Leg (1)
4	Core nodes* (7)	20	Left Foot (1)
5	Hip and Head (2)	21	Right Shoulder (1)
6	Hip and Neck (2)	22	Right Arm (1)
7	Hip and Spine (5)	23	Right Forearm (1)
8	Head and Neck (2)	24	Right Hand (1)
9	Head and Spine (5)	25	Left Shoulder (1)
10	Neck and Spine (5)	26	Left Arm (1)
11	Hip (1)	27	Left Forearm (1)
12	Head (1)	28	Left Hand (1)
13	Neck (1)	29	Spine 3 – close to Neck (1)
14	Spine (4)	30	Spine 2 (1)
15	Right Thigh (1)	31	Spine 1 (1)
16	Right Leg (1)	32	Spine 0 – close to Hip (1)

*Head, Neck, Spine, and Hip

RESULTS

Collect data and deploy five classifiers

A dataset containing 18,350 data points is collected from a subject performing the aforementioned motions with a 28-lb concrete block (Figure 4). Subject's motions are simultaneously videotaped to be used as the ground truth. Once the dataset is collected, five machine learning classifiers including logistic regression, k-nearest neighbors, multilayer perceptron, random forest, and support vector machine classifiers are deployed to recognize the motions from the dataset. Different sizes of the feature vector dimension based on the simulation scenarios are used to train and test the classifiers. 10-fold cross-validation is implemented to validate the classification performance. Hyper-parameters of each classifier are also tuned through the 10-fold cross-validation process.

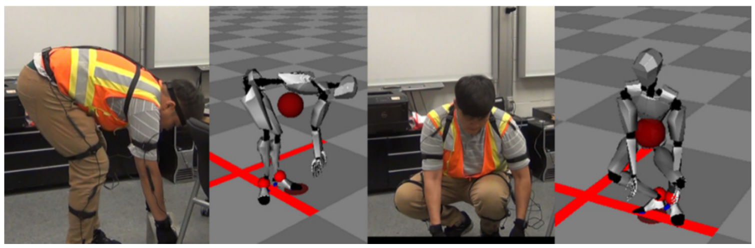

Figure 4. Bending and squatting motion examples.

Analyze classification performance

As shown in Figure 5, each classifier is evaluated in terms of accuracy which is the evaluation metric for the suitability of different sensor placement locations. As expected, the model using all nodes showed the best recognition performance. Among the five classifiers, random forest classifier showed the best performance in all cases. It is noteworthy that the model with two nodes located within a certain distance, such as combinations 5 and 6 as shown in Table 2, showed similar recognition performance compared to the model with all nodes. In the case of the single node scenarios, scenarios containing upper body nodes showed better recognition performance than the ones containing lower body nodes. While each classifier shows a different accuracy for the same scenario, varying the node combination for each classifier resulted in similar trends for accuracy.

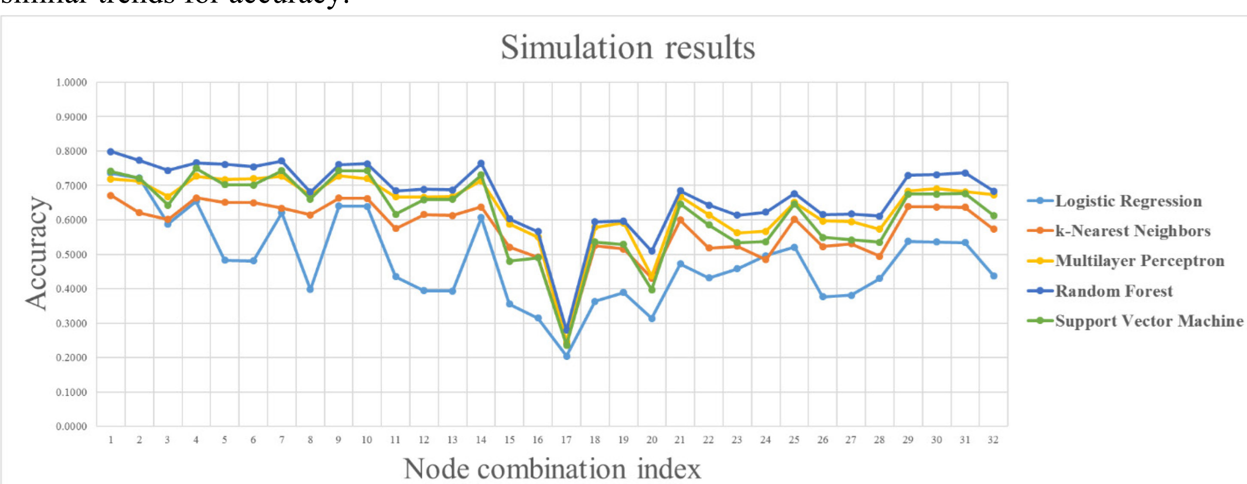

Figure 5. Simulation results with 32 node combinations

Table 2. Classification accuracy of Random Forest classifier

Combination	Selected nodes	Accuracy	Combination	Selected nodes	Accuracy
1	**All nodes**	**0.7983**	17	Right Foot	0.2791
2	Upper body	0.7729	18	Left Thigh	0.5941
3	Lower body	0.7440	19	Left Leg	0.5973
4	Core nodes	0.7658	20	Left Foot	0.5090
5	**Hip and Head**	**0.7617**	21	Right Shoulder	0.6844
6	**Hip and Neck**	**0.7544**	22	Right Arm	0.6431
7	Hip and Spine	0.7713	23	Right Forearm	0.6134
8	Head and Neck	0.6811	24	Right Hand	0.6222
9	Head and Spine	0.7606	25	Left Shoulder	0.6762
10	Neck and Spine	0.7635	26	Left Arm	0.6158
11	Hip	0.6849	27	Left Forearm	0.6171
12	Head	0.6893	28	Left Hand	0.6110
13	Neck	0.6871	29	Spine 3	0.7298
14	Spine	0.7639	30	Spine 2	0.7310
15	Right Thigh	0.6032	31	Spine 1	0.7365
16	Right Leg	0.5658	32	Spine 0	0.6833

CONCLUSION

This study investigated the impacts of IMU sensor locations on the motion recognition performance through a simulation. A set of 17 wearable IMU sensors was used to collect the motion data from the entire body. Five machine learning classifiers were deployed to evaluate their recognition performance based on the motion sensor locations. Comparing to the best performance node combination which includes all 17 nodes, the node combinations containing two nodes located in a certain distance, such as neck and hip or head and hip, showed similar motion recognition performance. It is an important finding in this study that fewer motion sensors can show a similar performance with the fully loaded sensors if their locations are selected with the understanding of their effectiveness based on the locations. Thus, a motion sensor-based system can effectively recognize worker's various motions with the reduced number of motion sensors, thus decreasing the implementation cost.

Future study will focus on several issues. First of all, the training dataset needs to be further collected from various subjects to improve the generality of the dataset. Since human motions vary depending on individual working behaviors, the dataset from the various subjects is essential to better represent the motions of construction workers. Second, the formation of the feature vector will be further investigated to achieve a better recognition accuracy. Raw data of IMU sensors is rarely used as input of existing motion recognition system. Instead, statistical features are extracted from the raw data and utilized as the input. These features may affect the recognition performance, and evaluation of the sensor location may also be changed.

REFERENCES

Akhavian, R., and Behzadan, A. H. (2016). "Productivity Analysis of Construction Worker Activities Using Smartphone Sensors." *Proceedings of 16th International Conference on Computing in Civil and Building Engineering*, 1067–1074.

Bayat, A., Pomplun, M., and Tran, D. A. (2014). "A Study on Human Activity Recognition Using Accelerometer Data from Smartphones." *Procedia Computer Science*, Elsevier, 34, 450–457.

Cho, Y. K., Ma, S., Ueda, J., and Kim, K. (2018). "A Wearable Robotic Exoskeleton for Construction Worker's Safety and Health." *2018 ASCE Construction Research Congress (CRC)*, New Orleans, LA, 19–28.

Escorcia, V., Dávila, M. A., Golparvar-Fard, M., and Niebles, J. C. (2012). "Automated Vision-based Recognition of Construction Worker Actions for Building Interior Construction Operations Using RGBD Cameras." *Proceedings of Construction Research Congress 2012*, 879–888.

Han, S., and Lee, S. (2013). "A vision-based motion capture and recognition framework for behavior-based safety management." *Automation in Construction*, Elsevier, 35, 131–141.

Kim, H., Paik, J., Park, J., and Park, H. (2016). "Motion Estimation-based Human Fall Detection for Visual Surveillance Low-light image restoration method View project Motion Estimation–based Human Fall Detection for Visual Surveillance." *IEIE Transactions on Smart Processing and Computing*, 5(5).

Kwapisz, J. R., Weiss, G. M., and Moore, S. A. (2011). "Activity Recognition using Cell Phone Accelerometers." *Proceedings of ACM SigKDD Explorations Newsletter 2011*, 74–82.

Michel, D., Qammaz, A., and Argyros, A. A. (2017). "Markerless 3D Human Pose Estimation and Tracking based on RGBD Cameras: an Experimental Evaluation." *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, Island of Rhodes, Greece, 115–122.

Nath, N. D., Akhavian, R., and Behzadan, A. H. (2017). "Ergonomic analysis of construction worker's body postures using wearable mobile sensors." *Applied Ergonomics*, Elsevier, 62, 107–117.

Peddi, A., Huan, L., Bai, Y., and Kim, S. (2009). "Development of Human Pose Analyzing Algorithms for the Determination of Construction Productivity in Real-Time." *Construction Research Congress 2009*, American Society of Civil Engineers, Reston, VA, 11–20.

Valero, E., Sivanathan, A., Bosché, F., and Abdel-Wahab, M. (2017). "Analysis of construction trade worker body motions using a wearable and wireless motion sensor network." *Automation in Construction*, Elsevier, 83, 48–55.

Yang, J. (2009). "Toward Physical Activity Diary: Motion Recognition Using Simple Acceleration Features with Mobile Phones." *Proceedings of the 1st international workshop on interactive multimedia for consumer electronics*, 1–10.

Yang, J., Shi, Z., and Wu, Z. (2016). "Vision-based action recognition of construction workers using dense trajectories." *Advanced Engineering Informatics*, Elsevier, 30(3), 327–336.

Yu, Y., Guo, H., Ding, Q., Li, H., and Skitmore, M. (2017). "An experimental study of real-time identification of construction workers' unsafe behaviors." *Automation in Construction*, Elsevier, 82, 193–206.