# AUTOMATED CONSTRUCTION WORKER PRODUCTIVITY ASSESSMENT USING SENSOR FUSION APPROACHES

GONG, YUE

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

The Department of Building and Real Estate

# Automated Construction Worker Productivity Assessment Using Sensor Fusion Approaches

GONG, Yue

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy

July 2023

# CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

______________________ (Signed)

_____GONG, Yue_____(Name of student)

# ABSTRACT

The construction industry is among the most labor-intensive sectors, with many tasks dependent on manual labor. This heavy reliance on a manual workforce presents challenges in managing ongoing projects, which is recognized as a fundamental cause of low productivity in construction. Therefore, identifying the worker-related root causes through continuous monitoring and assessment of activities is essential to addressing the low productivity issue. However, continuous activity monitoring for productivity analysis at the construction sites is challenging due to the environmental complexity and the dynamic nature of construction projects. Although observational methods are widely used to collect activity-related data, such as the locations of tasks performed and types of activities, these methods face criticism for their time-consuming and manual nature of data collection. For efficient monitoring of individual workers, previous research efforts have argued the need for automated approaches for field data collection by using sensing technologies, including cameras and wearable sensors.

Previous studies have proposed various sensor-based approaches for monitoring workers' activities. Nonetheless, significant challenges persist in identifying productivity issues. Firstly, most research in sensor-based activity monitoring categorizes activities based on repetitive tasks. However, the unstandardized nature of construction work means these predefined work taxonomies are not universally applicable, failing to recognize different working contexts essential for identifying core productivity problems. Secondly, existing sensor-based methods have primarily been validated in controlled environments, leaving the efficacy of these approaches for long-term, continuous activity data collection untested in field conditions. Thirdly, existing studies often depend on a single sensor data source, demonstrating acceptable accuracy in detecting various construction activities. However, each sensor has inherent

strengths and weaknesses, and reliance on a single data source could result in significant errors, particularly in challenging environments.

The current study aims to develop a comprehensive sensor-fusion-based automated activity assessment framework to identify potential worker-related productivity issues. The designed framework involves continuously collecting activity data using multi-modal sensors, including Bluetooth Low Energy (BLE) beacons for location tracking and accelerometers and cameras for activity monitoring. Specifically, this study established three objectives to address the research challenges outlined in the previous paragraph. The first objective is to design a refined taxonomy for construction activities, enhancing worker monitoring accuracy with work context information. The second objective is to assess the effectiveness of BLE beacon-based location tracking and accelerometer-based activity monitoring in diverse field settings. The third objective is to develop and evaluate a sensor fusion method that combines accelerometer and video data to improve activity recognition robustness in construction environments. The proposed worker activity assessment framework is expected to collect activity and location information from construction workers in real-time, aiding in better understanding individual worker-level productivity issues and determining the most suitable intervention strategies to improve construction productivity.

# ACKNOWLEDGMENTS

This research journey, filled with insights, discoveries, and growth, owes its completion to several people whose contributions were invaluable. Foremost, a debt of gratitude is extended to Ir. Dr. JoonOh Seo, my Chief Supervisor. Completing this research study owes a lot to his unwavering support, encouragement, and wealth of knowledge. His relentless guidance and sagacious counsel enriched this thesis and honed my research skills and academic perspective.

The contribution of my research colleagues and faculty members at the Department of Building and Real Estate of the Hong Kong Polytechnic University (PolyU) is greatly acknowledged. The assistance and cooperation extended during this study are sincerely appreciated by all PolyU's staff, faculty members, and research colleagues.

Expressing gratitude to Mrs. Dai Dai seems daunting as words are insufficient. Her unwavering faith, ceaseless support, and endless love provided a bedrock of strength during this PhD journey. Her resolute spirit and unconditional love remained the enduring source of motivation and inspiration. To my parents, who love and encourage me throughout this journey, it is hard to express the depth of my gratitude. Their relentless faith in my capabilities and their support propelled this research voyage to its destination.

To every person who contributed to this journey, either directly or indirectly, the value of their support is beyond words. This work stands as a testament to their invaluable contributions.

# TABLE OF CONTENTS

## CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AOA | Angle of Arrival |
| BBA | Basic Belief Assignment |
| BIM | Building Information Model |
| BLE | Bluetooth Low Energy |
| BML | Biological Motion Library |
| Bel | Belief Function |
| BiLSTM | Bidirectional Long Short-Term Memory |
| CI | Cell Identity |
| CII | Construction Industry Institute |
| CLP | Construction Labor Productivity |
| CNN | Convolutional Neural Network |
| CRC | Collaborative Representation Classifier |
| DAI-DAO | Data In–Data Out |
| DAI-FEO | Data In–Feature Out |
| DBN | Deep Belief Network |
| DEI-DEO | Decision In–Decision Out |
| DR | Drilling |
| DS | Dempster-Shafer |
| DST | Dempster-Shafer Theory |

| | |
|---|---|
| DoD | US Department of Defense |
| ECG | Electrocardiogram |
| EKF | Extended Kalman Filter |
| F1-WDS | F1-Weighted Dempster-Shafer |
| FEI-DEO | Feature In–Decision Out |
| FEI-FEO | Feature In–Feature Out |
| FME | Fuzzy Matter Element |
| FN | False Negative |
| FP | False Positive |
| FPS | Frames Per Second |
| GPS | Global Positioning Systems |
| HAR | Human Activity Recognition |
| HCI | Host Controller Interface |
| HMM | Hidden Markov Model |
| HR | Hammering |
| ID | Idling |
| IMU | Inertial Measurement Unit |
| IoT | Internet of Things |
| JDL | Joint Directors of Laboratories |
| KF | Kalman Filter |

| | |
|---|---|
| KNN | K-Nearest Neighbor |
| LAAS | Laboratory Analysis Architecture Systems |
| LB | Lifting Brick |
| LOOCV | Leave-One-Out Cross-Validation |
| LOS | Line-Of-Sight |
| LOSOCV | Leave-One-Subject-Out Cross-Validation |
| LR | Lifting Rebar |
| LSTM | Long Short-Term Memory |
| MHAD | Multimodal Human Action Database |
| MHEALTH | Mobile Health |
| MR | Measuring Rebar |
| NN | Neural Network |
| PF | Particle Filter |
| PL | Plausibility |
| R-FCN | Region-based Fully Convolutional Networks |
| RFID | Radio Frequency Identification |
| RNN | Recurrent Neural Networks |
| RSSI | Received Signal Strength Indicator |
| RTLS | Real-Time Location Estimation System |
| Recall-TopkWDS | Recall-Topk Weighted Dempster-Shafer method |

| | |
|---|---|
| ResNets | Residual Networks |
| Rx power | Received power |
| SVM | Support Vector Machine |
| TDOA | Time Difference of Arrival |
| TL | Traveling |
| TOA | Time of Arrival |
| TP | True Positive |
| TR | Tying Rebar |
| TWDS | Thresholding Weighted Dempster-Shafer |
| TopkWDS | Topk Weighted Dempster-Shafer |
| UUID | Universally Unique Identifier |
| UWB | Ultra-Wideband |
| WDS | Weighted Dempster–Shafer |
| WDST | Weighted Dempster–Shafer Theory |

# CHAPTER 1 INTRODUCTION

Construction is one of the most labor-intensive industries, and many construction tasks still rely on the manual workforce (Ng & Tang, 2010). High dependency on a manual workforce has been recognized as one of the fundamental causes of low productivity in construction (Jarkas, 2010). Task-level activity analysis has been utilized in the construction domain to identify the root causes of low productivity by continuously monitoring construction activities (Gouett et al., 2011). Notably, it quantifies the time spent on diverse activities and categorizes them based on their contribution to productivity (e.g., productive or non-productive), aiming to identify problematic operations requiring immediate intervention.

Recently, sensing technologies have shown potential for automatically collecting activity data (e.g., activity category and time expenditure) and location information. Collecting field data on workers' activities and locations using sensing technologies enables construction practitioners to understand the current status of diverse construction operations from a productivity perspective and identify causes of low productivity at the task-level. Previous research efforts have concentrated on various sensor-based approaches to automatically collect workers' activity and location data, utilizing machine learning/deep learning algorithms for action recognition and location tracking. Such automated action recognition frameworks employing machine learning-based classification have been widely used, demonstrating their potential to replace human observations with wearable sensors or cameras for continuous activity measurement without interfering with ongoing work (Hwang & Lee, 2017).

In construction activities, each type necessitates varied movements from the workers' bodies and joints, generating distinctive signal patterns (e.g., acceleration signals), which are able to be identified by action recognition classifiers. These models are trained using machine learning and deep learning algorithms to discern unique patterns from raw data, classifying different construction activities. Therefore, based on time-series acceleration or video data, action recognition automatically measures time spent on specific tasks within construction projects. Researchers in the field have explored the reliability and validity of automated activity recognition through acceleration data gathered in laboratory or construction sites, highlighting its significant potential for activity analysis (Akhavian & Behzadan, 2016; Bangaru et al., 2021b; Cheng et al., 2013; Joshua & Varghese, 2014; Kwapisz et al., 2011; Luís Sanhudo et al., 2021; Weiss et al., 2016).

Meanwhile, various entities, such as workers, equipment, and materials, are involved in construction operations. Monitoring and determining the location of construction entities is essential for various applications within construction sites, including resource optimization and progress monitoring (Dzeng et al., 2014). In particular, the location information acts as supplemental information to the activity recognition, resulting in higher classification credibility (Cheng et al., 2013). In the meantime, knowing work areas with low productivity allows us to understand potential issues related to low productivity. For example, low productivity would be affected by an individual worker's performance and operational factors such as lack of proper resources (e.g., workers, materials, and equipment). As a result, locating specific work areas with low productivity enables practitioners to immediately solve existing operational problems that could lead to low productivity. However, considering the large number of related entities working at construction sites, traditional localization with manual observation is labor-intensive and error-prone (Zhang et al., 2013), making the automated approach essential in tracking construction entities. Among various wireless technologies (e.g., Radio Frequency Identification (RFID), Global Positioning Systems (GPS), and Ultra-Wideband (UWB)) for tracking and locating construction entities, Bluetooth Low Energy (BLE) beacons have shown comparative advantages of 1) a low amount of infrastructure setting (J. Zhao et al., 2019), 2) flexible installation (Urano et al., 2017), 3) accessible to scalable both indoor and outdoor (Ng et al., 2020), and 4) cost-effective (Park et al., 2017). For example, unlike UWB, which requires a continuous power supply, battery-powered BLE beacons are more flexible to deploy in fast-changing environments (e.g., construction sites) (Khoury & Kamat, 2009). In contrast to wireless technologies, such as RFID and magnetic field, which need time-consuming calibration, the BLE beacon is capable of calibrating easily, therefore minimizing the infrastructure requirements (Park, Marks, et al., 2016). Due to these advantages and unique features of BLE beacons, previous studies have applied beacon-based location tracking of diverse construction entities, including construction workers (Park et al., 2017), resources (Shen et al., 2008), and vehicles (Lu et al., 2007).

Despite the potential of sensing technologies for task-level activity analysis, there are several remaining challenges in terms of 1) the interpretability of data obtained from sensing technologies, 2) the accuracy and reliability of sensor-based approaches, and 3) the sensor deployment at construction sites. First of all, as machine learning algorithms for action recognition deal with multiclass classification problems, how to define actions would significantly impact how to recognize productivity issues and whether machine learning algorithms can successfully learn unique data patterns according to activities. In the construction domain, the action categories tend to be determined on the basis of representative activities of construction tasks that are the most repeatedly performed. However, these activities may not provide sufficient information in the context of construction tasks that would be needed to understand productivity issues. For example, the delay of construction operations could occur due to a longer time for installing materials due to a lack of skilled carpenters, long material lead time caused by poorly optimized workspaces, or idling time caused by waiting for forms to be delivered to working areas. Even though these causes cannot be directly captured from action recognition, the actions should be precisely defined enough to implicitly identify these various work contexts associated with potential causes of low productivity. Also, the actions defined for action recognition algorithms often lead to confusion among different activities because of a lack of consideration of body movements that will directly affect sensor data patterns. Considering the non-standardized nature of field operations, the action recognition algorithms frequently suffer from noisy actions (e.g., actions that are unclearly predefined and labeled or transitional actions). These issues will be more remarkable in field operations where diverse activities are being performed in a continuous manner. In this regard, there is a solid need to re-design work taxonomy for action recognition algorithms by considering both work contexts and the distinguishability of data.

Secondly, relying on a single source of sensors to obtain field data for activity analysis would be risky, considering the multi-dimensionality of productivity issues and uncertainty caused by each sensor-based approach. This has led to the need for sensor-fusion approaches. As described above, activity analysis needs worker activity and location data. Thus, sensor fusion methods are needed to have the advantages of multiple sensing modalities, such as action recognition using wearable sensor- or vision-based approaches and location tracking methods such as BLE Beacons. From the activity analysis point of view, multi-modalities of field data (e.g., types and locations of activities) would help understand productivity issues better. Also, multi-sensor modality for a specific type of field data, such as types of activities, can reduce the uncertainty when different data sources (e.g., accelerations and images) are used individually (Lahat et al., 2015). For example, acceleration signals from a wearable sensor such as a wristband have shown good accuracy for classifying diverse construction activities but significant confusion between hand-dominant activities with similar upper-arm movements (Ryu et al., 2019). Instead, vision-based action recognition would be able to more accurately recognize diverse hand-dominant activities as images can capture whole-body moments. In this regard, the fusion of visual and non-visual modalities has been widely explored for human action recognition, showing better classification performance. However, most fusion methods for visual and non-visual modalities assume that both data are always available, and thus, machine learning algorithms can co-learn with two different modalities. Considering that images from construction sites frequently suffer from occlusions, it is expected to have only acceleration data at specific time frames, and thus, co-learning approaches are not possible. Also, during data fusion, balancing information from different sources is very important as each data modality may have different levels of confidence and reliability according to classes (Lahat et al., 2015). In this regard, the fusion approach should successfully work even without one modality of data, such as images, and maximize the complementarity of multimodality from multiple data sources.

Lastly, consideration of the dynamic nature of construction sites is vital in sensor data collection. Despite wearable sensors not being significantly impacted by working environments, deploying cameras and BLE Beacons requires careful consideration of site conditions. Specifically, signals from BLE Beacons are subjected to dynamic influences by various factors, such as distances, site layouts, signal propagation paths, and other environmental variables. This differs from cameras where site coverage can be visually assessed, thus introducing an additional layer of complexity in data collection.  The typical scenario of beacon-based tracking and localization is based on the distance measure (i.e., Rx power level approach) between the beacon and the receiver (e.g., smartphone) by using the characteristics of beacon signals that the signal strength would gradually decrease during propagation (Subhan et al., 2011). By using the estimated distances from multiple beacons (at least three), the receiver's position can be determined through trilateration methods (Elnahrawy et al., 2004; Han et al., 2007). However, the distance estimation is not always stable because the received signal tends to fluctuate as it is affected by environmental factors such as temperature and humidity (Amir Guidara et al., 2018). Also, the designed bandwidth of BLE technology does not allow the signal to penetrate obstacles like walls. Therefore, the signal received is the combination of signals from multiple paths, including directly received signals or signals reflected by walls, which could lead to inaccurate distance estimation (Faragher & Harle, 2014). This issue would be more significant, especially when deploying multiple beacons in the same area (Mackey et al., 2018). To mitigate signal frustration, previous studies have proposed and tested mathematical approaches to filtering out noisy signals, such as the Bayes filter, Kalman Filter (KF), Extended Kalman Filter (EKF), and Particle Filter (PF) (Xu et al., 2021). However, most studies mainly focused on signal noise and fluctuation during signal propagation without fully considering the impact of diverse environmental conditions on beacon signals.

To address these challenges, this study proposes automated task-level activity analysis by using sensor fusion approaches, in particular, focusing on the following research objectives:

Objective 1. Design a hierarchical work taxonomy for automated activity analysis.

Objective 2. Validate the feasibility of the proposed work taxonomy for automated action recognition using field data.

Objective 3. Develop a novel sensor fusion approach for action recognition by using both image and acceleration data, considering the independence and complementarity of multi-modality.

Objective 4. Evaluate a BLE beacon-based localization approach at various site conditions.

As depicted in Figure 1-1, the study begins with designing a work taxonomy that is comprehensive and universally applicable to construction tasks (Objective 1). The proposed taxonomy focuses on two essential aspects: 1) the potential productivity contribution of activities and 2) the existence of distinctive body movements. The first aspect facilitates valuable information extraction in the context of activity contribution, thus enhancing the interpretability of data analysis. Conversely, the second aspect aids in distinguishing motions through inherent data characteristics, thereby mitigating classification errors from action recognition algorithms. Next, the feasibility of this proposed taxonomy was validated for automated action recognition utilizing field experiments (Objective 2). The field data were collected in an uncontrolled manner from 18 construction workers at two construction sites in Hong Kong. Over two months, acceleration data was gathered during concrete work tasks such as formwork and rebar installation using an inertial measurement unit (IMU) embedded in a smartwatch (e.g., Apple Watch). Subsequently, the acquired data was labeled according to the proposed work taxonomy. The validation process incorporated both traditional feature-based machine learning and advanced deep learning algorithms, all for acceleration-based action recognition. The field experiment was designed to evaluate the validity of the proposed taxonomy and the classification accuracy and reliability of the acceleration-based activity recognition model.

Moreover, the present study formulated an action recognition framework that combines vision and acceleration-based models using decision-level fusion approaches and confirmed the sensor fusion’s complementarity through laboratory experiments (Objective 3). Eight activities from the taxonomy designed in Objective 1 were selected in the lab test. Three participants were employed to perform each activity five times sequentially. Data were concurrently recorded via an Apple Watch used in the previous field test (Objective 2) and three smartphones (iPhone) to capture acceleration and three angles’ video data. Initial experiments assessed the taxonomy’s effectiveness in a video-based activity recognition model with laboratory-collected data. Subsequently, acceleration and video data were utilized to train the model separately. The proposed decision-level fusion approaches were then employed to integrate preliminary estimates from the models trained from sole-sensor sources. Performance comparison would affirm the fusion network’s efficacy and thereby underscore the complementary benefits of using the fusion approach in construction activity recognition.

Lastly, the current study designed various environmental conditions and scenarios for an in-depth analysis of BLE beacon signals, thereby understanding the deployment principle of BLE beacon for the localization application in the construction site (Objective 4). The proposed approach involved varying the beacon installation height, signal receiver position, and the indoor environment's geometry. Field tests were conducted at a construction site, leveraging commercial beacon devices. Data were collected using an iOS application and included twelve independent trials featuring two different testbeds and 24 unique beacons. The dataset spanned a month, providing substantial evidence for signal strength analysis and contributing to understanding and improving sensor deployment in varying construction site conditions.

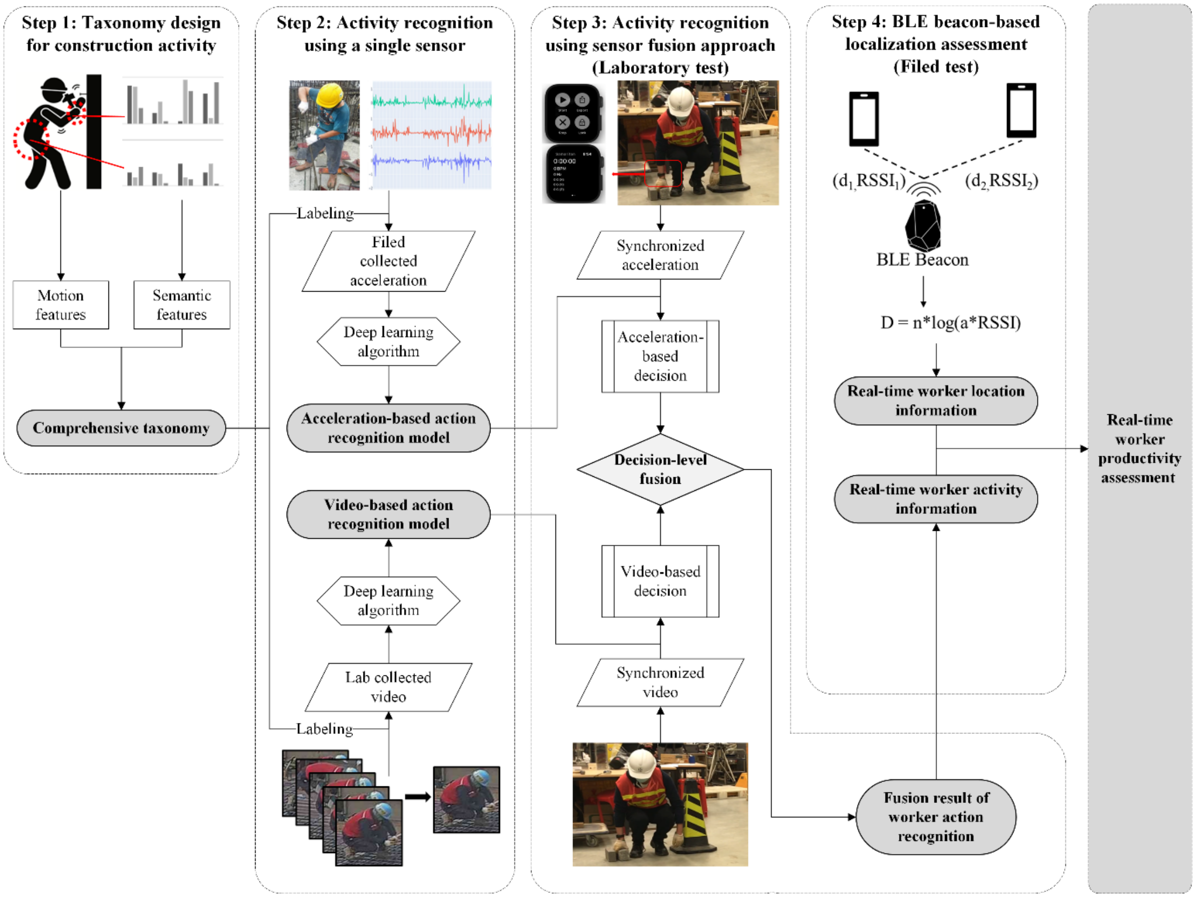

Step 1: Taxonomy design for construction activity
Motion features
Semantic features
Comprehensive taxonomy
Step 2: Activity recognition using a single sensor
Labeling
Filed collected acceleration
Deep learning algorithm
Acceleration-based action recognition model
Video-based action recognition model
Deep learning algorithm
Lab collected video
Labeling
Step 3: Activity recognition using sensor fusion approach (Laboratory test)
Synchronized acceleration
Acceleration-based decision
Decision-level fusion
Video-based decision
Synchronized video
Step 4: BLE beacon-based localization assessment (Filed test)
(d1,RSSI1)
(d2,RSSI2)
BLE Beacon
D = n*log(a*RSSI)
Real-time worker location information
Real-time worker activity information
Fusion result of worker action recognition
Real-time worker productivity assessment

Figure 1-1 An overview of the research framework

# CHAPTER 2 TAXONOMY DESIGN OF TASK-LEVEL ACTIVITY ANALYSIS[1]

## 2.1 Task-Level Productivity Assessment

Productivity stands as a crucial determinant of project success in the construction industry, with significant influence over cost, time, and quality (Nasirzadeh & Nojedehi, 2013). Despite various interpretations, productivity is generally understood as the construction process's ratio of outputs to inputs. In the construction industry, the output is quantifiable in diverse units, including components assembled, square meters, or cubic meters of material installed. Conversely, labor is the primary input of productivity measurement because of the labor-intensive nature of construction projects. In this regard, Construction Labor Productivity (CLP), defined as the number of work hours required for unit output, is the widely preferred metric for construction productivity assessment (Yi & Chan, 2014). Within the construction industry, the evaluation of labor productivity occurs at three distinct levels: task-, project-, and industry-level, each signifying an increasing level of complexity. The task-level assessment, which focuses on individual construction activities, is widely employed within the industry. For instance, the Construction Industry Institute (CII) utilizes such metric as a benchmark for productivity estimation (Chapman et al., 2010). Compared to the project- or industry-level, task-level assessment functions as a single-factor productivity measure with controlled variables. As a result, the root cause of low productivity can be more efficiently identified by

[1] This chapter is partially based on a published study and being reproduced with the permission of Elsevier.

**Gong, Y.**, Yang, K., Seo, J., & Lee, J. G. (2022). Wearable acceleration-based action recognition for long-term and continuous activity analysis in construction site. *Journal of Building Engineering*, *52*, 104448.

focusing on individual task tracks rather than the collective trajectory derived from numerous task streams within a project. Furthermore, tasks are a construction project's foundational units associated with minimal time investment. As a consequence, task-level productivity deficiencies can be identified closest to the time when the issue arises. Such an instant way of exposing productivity issues allows immediate intervention to resolve the problem.

Various productivity assessment methods have evolved, including Productivity Measuring Methods (PMMs) and Productivity Improving Methods (PIMs) (Adrian & Boyer, 1976). Despite their utility, traditional methods may exhibit drawbacks, such as labor intensiveness and failures to deliver accurate and timely responses. Techniques such as activity analysis have earned widespread adoption in productivity assessment to overcome such limitations. As a developed form of the sampling method, this technique focuses on the identification of activity categories while simultaneously tracking the time spent on each activity pattern, which aids in the estimation of task-level labor productivity. When applied in the construction industry, it enables continuous assessment, serving as productivity feedback that highlights tasks and workers requiring further improvements.(Jacobsen et al., 2023).

The task-level activity analysis serves as a vital tool for improving efficiencies within the construction industry. In the assessment procedure, the activity analysis method is employed to quantify labor productivity in the context of labor hours spent on physical tasks. Given the productivity estimation of target tasks, the productivity assessment facilitates promptly finding the exact task inherent to low-productivity issues by focusing on discrete tasks that comprise larger projects. As a result, interventions can be initiated to address these productivity challenges. Success in resolving such issues leads to improvements in the construction process. Despite the crucial importance of task-level activity analysis, its application presents considerable challenges. One notable difficulty lies in defining tasks accurately, reasonably, and consistently, which would ensure that measurements at the task level remain objective and comparable.

## 2.2 Literature Review of Activity Taxonomy Design

Addressing the challenge of applying task-level activity analysis in construction productivity assessment requires a robust task definition system, which requires a comprehensive taxonomy system explicitly designed for construction activities. A well-designed taxonomy would facilitate clearly labeling collected data, reducing confusion and ensuring every data pattern is accommodated within the labeling system. In the meantime, the well-structured activity taxonomy could facilitate a better understanding of data, thereby enriching the interpretability of the resulting analyses.

In order to find the principles of designing a robust activity taxonomy system, the current study investigated the work taxonomy employed in previous activity recognition research. The findings are summarized in Table 2-1, showing that action recognition typically classifies activities either as 1) movement-oriented tasks such as standing, walking, hammering, and screwing or 2) work context-oriented tasks encompassing actions related to masonry work like spreading mortar, fetching bricks, and filling joints. The use of body movement characteristics for task definition often leads to a more accurate activity classification, given that distinct movements create unique responses. This approach, however, may lack semantic properties due to the absence of work context information. Conversely, employing work context-oriented tasks as labels for action recognition can yield more intuitive knowledge when measuring work expenditures during activity analysis, thereby aiding in construction labor assessment and delay identification. Such a mode of activity recognition, though, can result in poor classification performance when the tasks being classified involve similar body movements. For example, formwork activities viewed from a work context perspective involve both assembling and stripping formwork, which shares the common task of hammering. Thus, an activity taxonomy for action classification must encompass both movement and work context to deliver high-performance classification results enriched with comprehensive information on construction activities.

Table 2-1 Activity taxonomy used in the Human Activity Recognition (HAR) domain

| **Taxonomy criteria** | **Activity category** | **Classification model** | **Classification accuracy** | **Data collection method** | **Research** |
|---|---|---|---|---|---|
| Motion | Basic task: Connecting, covering, cutting, digging, finishing, inspecting, measuring, placing, planning, positioning, spraying, spreading | - | - | Observation | Everett and Slocum (1994) |
| Motion | Walking, tying rebar guiding crane between activities | - | - | Automation (camera) | Buchholz et al. (2003) |
| Motion | Loading, pushing, unloading, returning, idling | Neural network, decision trees, K-Nearest Neighbor (KNN), logistic regression, and Support Vector Machine (SVM) | 87% to 97% (user-dependent) and 62% to 96% (user-independent) | Automation (smartphone) | Akhavian and Behzadan (2016) |
| Context | Work, material, travel, and idle | - | - | Automation (location sensor and accelerometer) | Cheng et al. (2013) |
| Context | Direct work, tools and materials, instructions and drawings, crane deliveries, minor contributory work, travel, idle, unexplained, waiting, no contact | - | - | Observation | Thomas and Daily (1983) |

| | | | | | |
|---|---|---|---|---|---|
| Context | Effective work, essential contributory work, ineffective work | Decision tree | 90.1% (ironwork) and 77.7% (carpentry) | Automation (IMU) | Joshua and Varghese (2014) |
| Context | Spreading mortar, laying blocks, adjusting blocks, removing mortar | KNN, multilayer perceptron, decision tree, and multiclass Support Vector Machine | 88.1% | Automation (IMU) | Ryu et al. (2019) |
| Motion | Sitting, lying down, walking, walking upstairs, walking downstairs, stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand | Deep Belief Network (DBN) | 89.6% | Automation (accelerometer) | Hassan et al. (2018) |
| Motion | Jogging, walking, upstairs, downstairs, sitting, standing | Convolutional Neural Network (CNN) | 97.6% | Automation (accelerometer) | Ignatov (2018) |
| Motion | Run, walk, still | CNN | 92.7% | Automation (accelerometer) | Lee et al. (2017) |
| Motion | Biological Motion Library (BML): knocking, lifting, throwing, walking. Multimodal Human Action Database (MHAD): jumping, jumping jacks, bending, punching, waving (two hands), waving (one hand), clapping, throwing, sit-down/stand-up, sit-down, stand-up | Recurrent Neural Networks (RNN) | 99% (BML) and 99% (MHAD) | Automation (magnetic induction sensor) | Golestani and Moghaddam (2020) |
| Motion and context | PAMAP2 Dataset: lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, ascend and descend stairs, watch TV, computer work, | Inception neural network, Recurrent Neural Network | 94.5% | Automation (accelerometer) | Xu et al. (2019) |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
|  | drive the car, fold laundry, clean house, play soccer |  |  |  |  |
| Motion | MHEALTH Dataset: standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bends forward, frontal elevation of arms, knees bending, cycling, jogging, running, jumping front and back | Neural networks with Simple Recurrent Units (SRUs) and Gated Recurrent Units (GRUs) | 99.6% | Automation (accelerometer and ECG) | Gumaei et al. (2019) |
| Motion | Standing, bending-up, bending, bending-down, squatting-up, squatting, squatting-down, walking, twisting, working overhead, kneeling-up, kneeling, kneeling-down, and using stairs | Long Short-Term Memory (LSTM) | 94.7% | Automation (IMU) | Kim and Cho (2020) |
| Motion | Adjusting jacks, carrying crossbars, carrying jacks, carrying scaffold plank, carrying scaffold frame, dragging scaffold plank, hammering, inserting jacks into scaffold frame, lifting scaffold plank from elbow to overhead, walking, wrenching, climbing, downstairs, climbing with tool bag, downstairs with tool bag | Neural network | 93.3% | Automation (accelerometer and ECG) | Bangaru et al. (2021a) |

## 2.3 Comprehensive Construction Activity Taxonomy Designed with a Hierarchical Structure

The prior section investigated that movement characteristics and work context are essential principles of designing a well-performed activity analysis taxonomy. Based on these principles, a comprehensive activity analysis taxonomy was designed to assess task-level construction productivity. The proposed comprehensive activity taxonomy consists of three hierarchical levels of activities to extract activity-related information and better understand the work context performed by a construction worker (Table 2-2).

Table 2-2 Construction activity taxonomy designed for worker's productivity assessment

| **Activity level** | Activity | | | | | | |
|---|---|---|---|---|---|---|---|
| **Level 1** | Idling | Work | | | | | |
| **Level 2** | Stationary (Ineffective) | Traveling (Supportive work) | | Material Installation (Effective work) | | | |
| **Level 3** | Standing or sitting | Transportation | Transferring materials and tools | Material preparation | Material connecting | Material placing | Supplement work |
| **Basic task** | Standing, and sitting | Horizontal, vertical, and inclined movement, jumping, striding, going upstairs or downstairs, climbing up or down a ladder | Carrying materials in horizontal, vertical, and inclined movement, carrying materials while going upstairs or downstairs and climbing ladders, dynamical wrist movement while traveling | Rebar work: cutting, bending<br>Formwork: cutting, measuring, and drawing | Rebar work: fixing, tying, installing stirrup<br>Formwork: screwing, drilling, knocking, removing nails | Rebar work: placing, adjusting, lifting<br>Formwork: Attaching, adjusting, and lifting formwork | Lifting materials and tools, squatting, standing up, rotating trunk, transition movement |

The first criterion for categorizing activities is whether the activity is relevant to the production process, and the activities are classified into "Idling" (e.g., standing and sitting) or "Work" at Level 1. As an offspring activity category of "Work," activities at Level 2 are defined in accordance with activity-related movements, depending on whether they involve hand-dominant or whole-body-dominant movements. As the acceleration data are collected from a smartwatch, the signals will be more dominantly affected by hand movements and less affected by whole-body movements. By classifying Level 2 activities into "Traveling," which involves horizontal whole-body movements, and "Material installation," which is associated with hand-dominant activities, the acceleration signals from the two activities can be more distinguishable. Such a difference is illustrated in Figure 2-1. Meanwhile, three activity categories at Level 2, which include "Stationary," "Traveling," and "Material installation," can provide information to evaluate the work efficiency of the operations to be monitored. For example, the longer time spent on "Material installation" may indicate that the operation will be more efficient for producing outputs. The activities at Level 3 focus more on understanding the work context that will help identify productivity inhibitors. For this purpose, "Traveling" is further classified into "Transportation" and "Transferring materials and tools" at Level 3, and "Material installation" is divided into four subactivities, including "Material preparation," "Material connecting," "Material placing," and "Supplement work." Detecting the problematic activities that can lead to inefficiency in activities at Level 2 is possible by further classifying activities at Level 3. However, as the activity categories at Level 3 are based on general work contexts, they can be applicable to any other construction operations that involve delivering and installing materials for specific building components. However, some activities, including intermittent or supportive activities for other activities at Level 3, are unclearly classified on the basis of work contexts. These activities are included in "Supplement work". Table 2-2 shows examples of the basic tasks that can be included in activities at Level 3 for rebar work and formwork that are operations to be tested in this study. For example, “Material preparation,” which refers to producing components for further operation, can include several basic tasks, such as cutting, bending, and drilling. “Material connecting” is the assembling tasks, including fixing, tiling, screwing, and knocking, and material placing represents the lifting and adjusting of associated components. “Supplement work” includes all supportive movements that occur during the installation process. The basic tasks can be used for a better categorization of the activity taxonomy in this study and for precisely recognizing the labeling procedure. Such segmentation of tasks for activities in Level 3 will help understand the context of activities but will also increase the uncertainty of an automated activity classification using a wearable sensor. Specifically, the classification of Level 3 activities is questionable due to the similarity and dissimilarity of acceleration signals from different activities. For instance, knocking and cutting movements are the offspring activities of “Material installation” that will generate cyclic acceleration data with repetitive hand movements. Consequently, distinguishing the Level 3 activities for “Material installation” solely by hand movements is difficult because each category of the operation comprises dynamic and complex hand movements. Transportation and transferring of materials/tools will have different hand movements. The hand will swing periodically in “Transportation” activities (e.g., walking) or sway (e.g., adjusting tool while walking) mildly (e.g., holding material steady while walking) in “Material or tool transferring” activities. These facts lead to this research investigating the activity classification performance with a proposed activity taxonomy (Table 2-2).

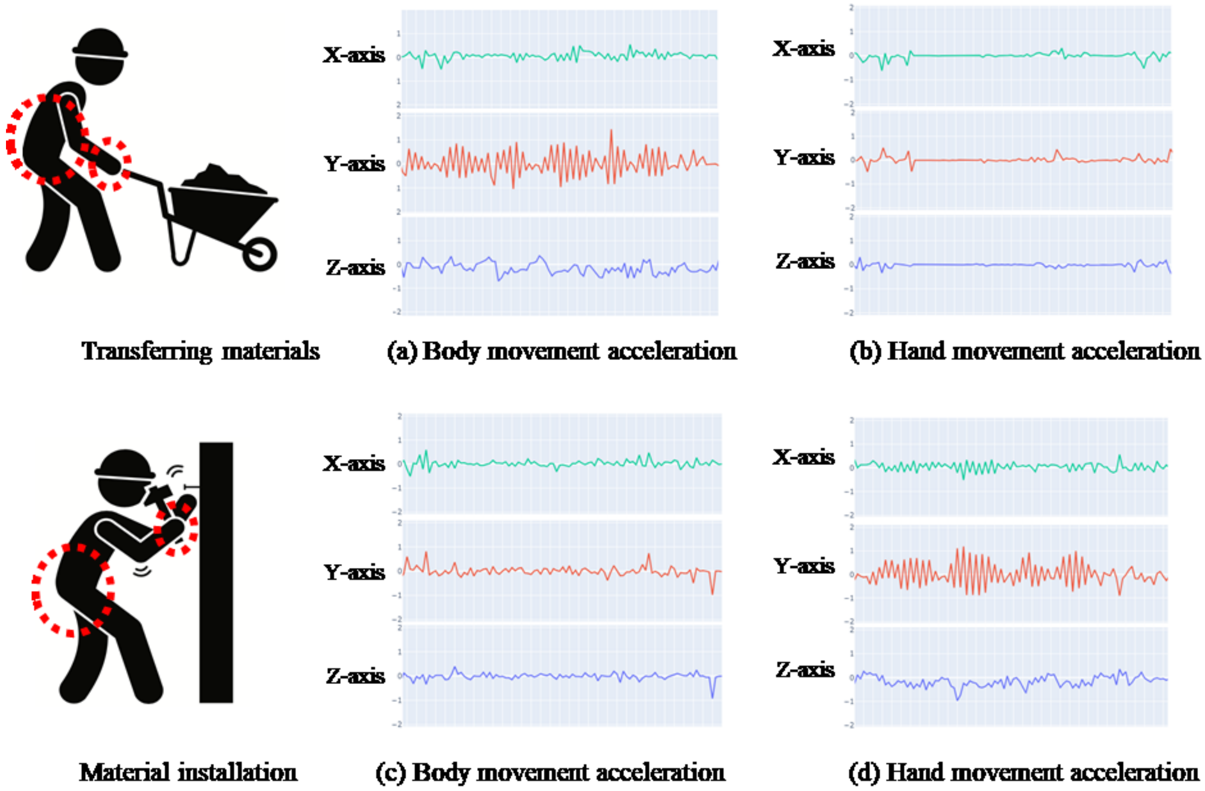


Figure 2-1 Acceleration signals of hand-dominant activity and body-dominant activity

# CHAPTER 3 FIELD VALIDATION OF ACCELERATION-BASED TASK-LEVEL WORKER PRODUCTIVITY ASSESSMENT[2]

## 3.1 Background

Recently, the construction sector witnessed the introduction of automated action recognition techniques using sensors facilitated by machine learning methods. The application has shown its potential to replace human observers for continuous activity measurement without interfering with ongoing work (Hwang & Lee, 2017). Despite its usefulness, a few challenges have been identified concerning its practical implementation in ongoing construction tasks. Nonetheless, a few challenges associated with its implementation in ongoing construction tasks have been highlighted. Specifically, the definition of activities poses a significant influence on the performance of automatic activity recognition.

In the construction domain, the action categories tend to be determined on the basis of representative activities of construction work that are the most repeatedly performed. However, confusion among different activities frequently occurs because of the lack of consideration of body movements that will directly affect the pattern of motion signals from body-attached sensors. Considering the nonstandardized nature of field operations, the action recognition algorithms frequently suffer from noisy actions (e.g., actions that are unclearly predefined and

[2] This chapter is based on a published study and being reproduced with the permission of Elsevier.

**Gong, Y.**, Yang, K., Seo, J., & Lee, J. G. (2022). Wearable acceleration-based action recognition for long-term and continuous activity analysis in construction site. *Journal of Building Engineering*, *52*, 104448.

labeled or transitional actions). These issues will be more remarkable in data that are continuously collected in unstructured settings, such as actual construction sites.

To address the limitations, the authors of this study propose a new work taxonomy that considers movement and work contexts (Table 2-2). This taxonomy aims to extract useful information for activity analysis and reduce classification errors from action recognition algorithms. The proposal of a comprehensive and universally applicable work taxonomy for construction tasks considers 1) the potential of activities to contribute to productivity and 2) the engagement of activities in unique body movements that may generate distinguishable acceleration signals, liable to serve as features for classification. However, the effectiveness of this proposed taxonomy in analyzing construction activity remains uncertain. Hence, this chapter seeks to validate its effectiveness by analyzing worker activity data collected from the construction field. The validation employed the acceleration sensor, which is widely used in the human activity recognition domain and is ideally suitable for construction demand. Diverse construction activities involve specific body movements of construction workers, and these movements create unique acceleration signals. Acceleration-based action recognition tries to automatically capture these unique patterns from the signals by using machine learning algorithms and classify diverse construction activities. As action recognition is performed on the basis of a set of time-series acceleration data, the classification results can be used to measure the time spent on specific activities in any construction tasks automatically. Several researchers in construction have examined the reliability and validity of automated activity recognition by using acceleration data collected in laboratory settings or construction sites and demonstrated its great potential for activity analysis (Akhavian & Behzadan, 2016; Bangaru et al., 2021b; Cheng et al., 2013; Joshua & Varghese, 2014; Kwapisz et al., 2011; Luís Sanhudo et al., 2021; Weiss et al., 2016).

In the validation, both traditional feature-based machine learning and deep learning algorithms for acceleration-based action recognition were used. In particular, acceleration data are collected from 18 construction workers from two construction sites in an uncontrolled manner by using an inertial measurement unit (IMU) embedded in a smartwatch (i.e., Apple Watch) during concrete work (e.g., formwork and rebar installation) for two months. The collected data are labeled in accordance with the proposed work taxonomy to evaluate the validity of the taxonomy and the classification performance by applying various machine learning algorithms. On the basis of the action classification results, the usefulness of the proposed work taxonomy and its appropriate level of detail are discussed. Future research directions to enhance the practicability of automated activity recognition and activity analysis in a construction workplace are explored.

## 3.2 Methodology

This research proposes a comprehensive activity taxonomy considering the characteristics of workers' movements and the work context that will serve as action labels for acceleration-based recognition algorithms and investigates the validity of the algorithms in practice by using continuously collected field data. Figure 3-1 illustrates the overall research framework. A comprehensive activity taxonomy aiming to effectively measure activities required for identifying productivity issues while minimizing possible confusion in action classification was proposed. For field validation, two local construction sites in Hong Kong were recruited, and continuous acceleration data during construction works (e.g., rebar and formwork) were collected by using an IMU-embedded smartwatch. The videos were simultaneously recorded by using a chest-mounted portable video camera for labeling activities. Machine learning-based classification algorithms were applied to the collected acceleration data to classify diverse activities that were defined on the basis of the proposed activity taxonomy. Thereby, the validity of the proposed activity taxonomy for action recognition and its applicability for productivity assessment were examined on the classification performance.

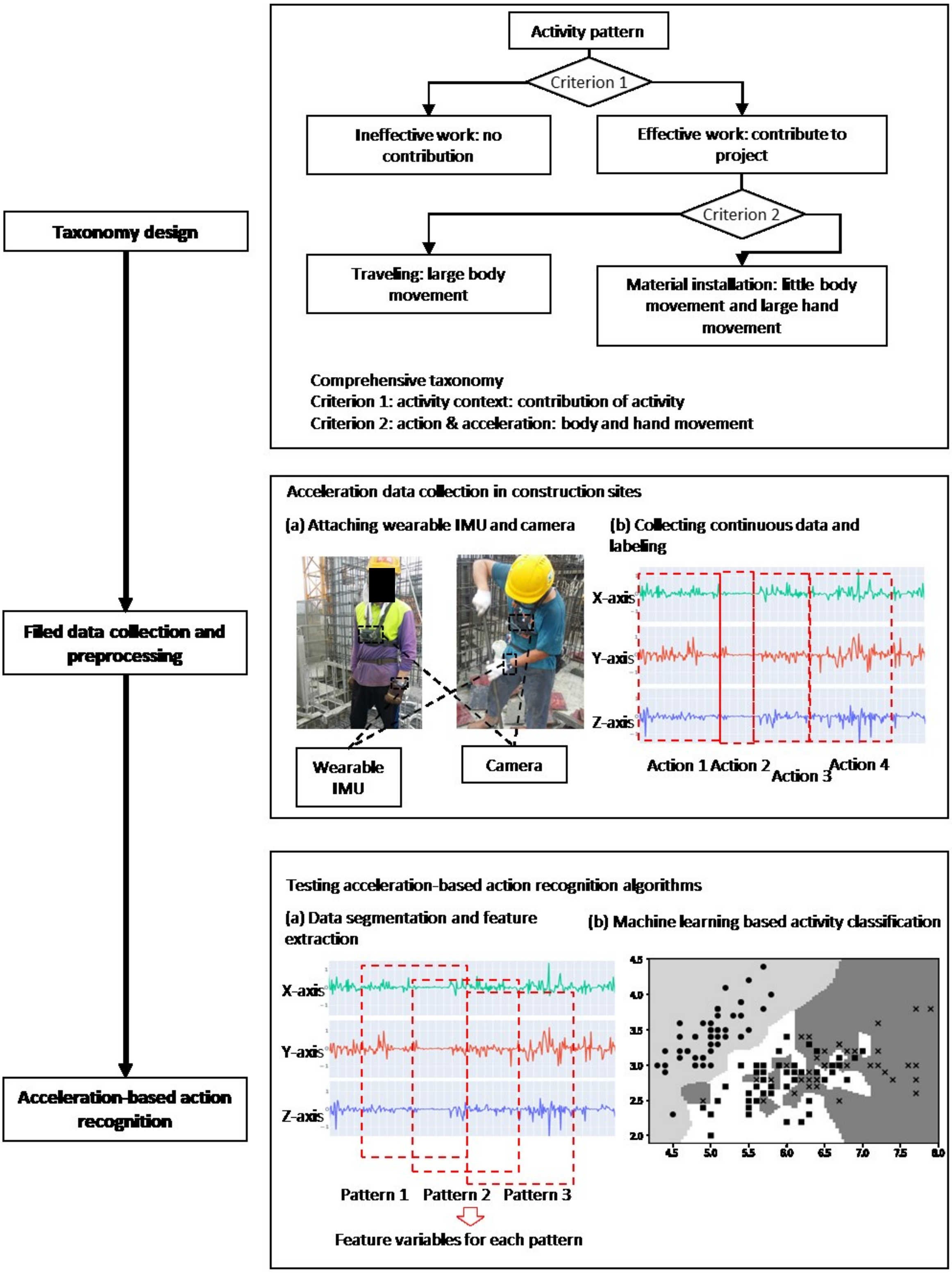

Taxonomy design
Filed data collection and preprocessing
Acceleration-based action recognition
Activity pattern
Criterion 1
Ineffective work: no contribution
Effective work: contribute to project
Criterion 2
Traveling: large body movement
Material installation: little body movement and large hand movement
Comprehensive taxonomy
Criterion 1: activity context: contribution of activity
Criterion 2: action & acceleration: body and hand movement
Acceleration data collection in construction sites
(a) Attaching wearable IMU and camera
(b) Collecting continuous data and labeling
X-axis
Y-axis
Z-axis
Wearable IMU
Camera
Action 1
Action 2
Action 3
Action 4
Testing acceleration-based action recognition algorithms
(a) Data segmentation and feature extraction
(b) Machine learning based activity classification
X-axis
Y-axis
Z-axis
Pattern 1
Pattern 2
Pattern 3
Feature variables for each pattern
4.5
4.0
3.5
3.0
2.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

Figure 3-1 Research framework of acceleration-based worker productivity assessment

### 3.2.1 Collecting and preprocessing of data

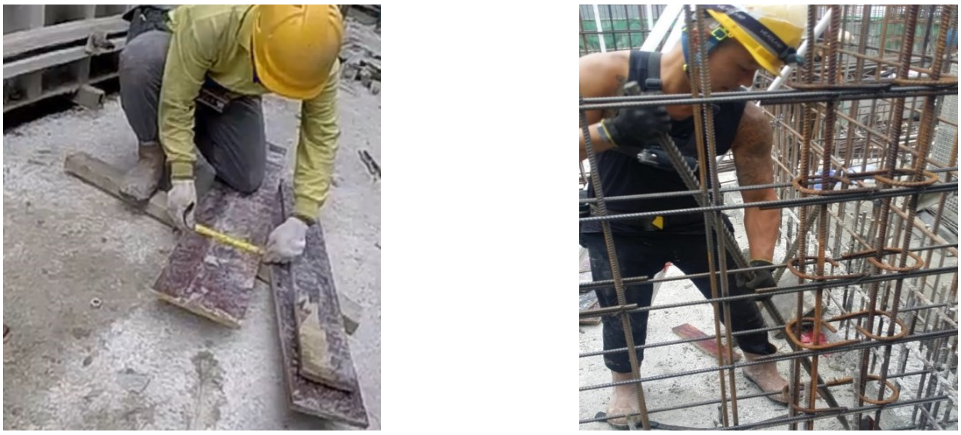


Figure 3-2 Site photos for data collection

Data collection was performed during formwork and rebar work (Figure 3-2) to study the validity of the proposed activity taxonomy and the performance of the acceleration-based activity recognition approach. Nineteen individual periods were involved in the data collection, and each period lasted a whole workday. A large-scale dataset that included 498 h of videos and 2.8 billion samples of acceleration data was constructed from 18 construction workers. Each participant was equipped with an Apple Watch embedded with a sensor in the dominant hand to record cumulatively 3D acceleration data through a self-developed watchOS app. The frequency of data collection was set to 100 Hz, indicating that the wearable sensor recorded 100 acceleration data sets for each second. A chest-mounted GoPro camera was used to record simultaneous hand movements for the data labeling. The videos were recorded at 30 FPS, allowing the ground truth of activity information to be captured and stored in a stable and durable manner. The data collection was conducted for two sessions per day (i.e., morning session and afternoon session), and each session lasted two hours or so.

The equipment was taken off during the lunch break because the device needed to be calibrated again before starting the afternoon session. The collected acceleration signals were labeled for each data point based on the researchers' observations on video recordings. Each video frame was labeled by using one of the activities defined at each level of the proposed activity taxonomy based on the observer's judgment. Corresponding acceleration signals were labeled by comparing time information for each data point. In some video scenes, workers' hand activities were unclearly captured. In this case, the activities were determined on the basis of the observations of overall sequences of activities. However, one of the challenges for data labeling is to judge the boundary of consecutive activities. The boundary was determined on the basis of the starting time of the following activity for consistent labeling. If there are significant transitions between two consecutive activities, then these transitional activities were labeled as "Supplement work" considering their work contexts. Unqualified data, such as those collected under suboptimal lighting conditions or data recorded during a break in the restroom, were omitted from subsequent processing to avoid possible confusion caused by bad judgment on ongoing activity.

### 3.2.2 Machine learning-based activity recognition

Traditional machine learning and deep learning algorithms were applied to test the applicability of the proposed activity taxonomy. A sliding window technique was applied when segmenting labeled acceleration signals into patterns of equal size because any human activity should last for a particular duration (Banos et al., 2014). In the current study, a 50% overlap was adopted to reduce the transition noise (Su et al., 2014). The length of the window was determined by considering the nature of the construction activity. On the basis of the experience of previous research (Ryu et al., 2019), this study tested multiple window lengths (i.e., 0.5, 1.0, 1.5, 2.0, 2.5, 3, 3.5, and 4.0 s) and determined the optimal window length in accordance with the classification accuracy. The activity labels of each segmented data were determined on the basis of the majority voting rule when data points with multiple activity labels were found within the window (Ballabio et al., 2019).

For classifiers of different activities, this study investigated traditional feature-based machine learning and deep learning approaches for performance comparison. As traditional machine learning classifieds, we selected three classifiers that had been widely applied for activity recognition, namely, 1) ensemble bagged trees (Dietterich, 2000), 2) support vector machine (Hsu & Lin, 2002), and 3) k-nearest neighbor (Sutton, 2012). The Classification Learner app in MATLAB (2019a, MathWorks) was utilized to train and test the models for identifying the best-performing classifier and corresponding hyperparameters, aiming to validate the feasibility of the proposed taxonomy. Typical features applied in activity recognition were time-domain features and frequency-domain features (Preece et al., 2009). Time-domain features interpret the statistical characteristics of motion signals, including but not limited to the mean, maximum, median, and variance of the signals (Figo et al., 2010). Specifically, this study used eight time-domain features that consist of mean value, minimum value, maximum value, range, standard deviation, kurtosis, correlation, and skewness of acceleration signals in the *X*, *Y*, and *Z* axis. Two frequency-domain features, energy and entropy, were used to capture the acceleration streams in terms of frequency, which evaluate action complexity in acceleration-based activity analysis (Ryu et al., 2019). Fast Fourier transform was applied to extract frequency-domain features from raw signals (Preece et al., 2009). This study tested deep learning algorithms that had the comparative benefits of eliminating the need for hand-crafted features and can save time and effort in the selection and optimization of features and the reduction of human bias (Krizhevsky et al., 2012). This study implemented a Bidirectional Long Short-Term Memory (BiLSTM), one of the deep learning algorithms known to provide reliable classification performance for acceleration-based action recognition (Yang et al., 2020). The designed architecture is shown in Figure 3-3.

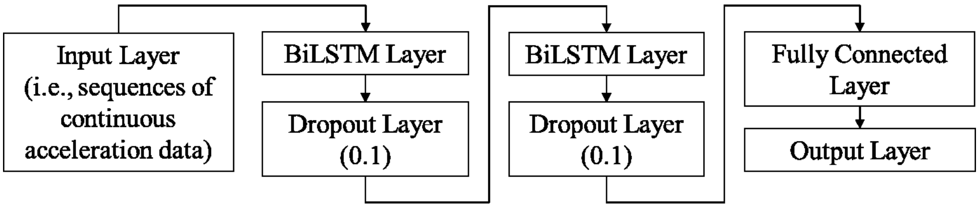


Figure 3-3 Architecture of the deep learning algorithm for acceleration-based activity recognition

Two types of cross-validation techniques were applied to evaluate the performance of classifiers for each level of activities: 1) leave-one-out cross-validation (LOOCV) and 2) Leave-One-Subject-Out cross-validation (LOSOCV). For leave-one-out cross-validation, the whole data set was randomly separated into five exclusive subsets of equal sizes. Each subset was utilized as testing data for each trial of validation, and the remaining datasets were used for training the machine learning models. The average prediction accuracy of the five validation tests was regarded as the classification performance of the designed algorithm, indicating the overall accuracy of the trained model (Refaeilzadeh et al., 2009). To investigate subject-to-subject variation, we conducted the LOSOCV, which selects one worker's data as testing data once at a time and the data from other workers for training the models (Berrar, 2019). The classification models were trained and tested using different activity data levels (Levels 1, 2, and 3) to examine whether the classification results at each level will be accurate and reliable for understanding productivity issues during construction operations. The action classification results at each work taxonomy level are presented using the confusion matrices, where each row represents actual classes, and each column corresponds to predicted classes (Mantyjarvi et al., 2001). In particular, recall quantifies the fraction of positive observations correctly predicted, while precision calculates the ratio of correct predictions that are indeed positive (Davis & Goadrich, 2006). These metrics can be calculated using Equations (3-1) and (3-2), where TP denotes True Positive, FP stands for False Positive, and FN represents False

Negative. The F1 score provides a balance between these two metrics, especially when there is an uneven class distribution (Equation (3-3)).

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{3-1}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3-2}$$

$$\mathrm{F1} = \frac{2\times(\mathrm{Precision}\times\mathrm{Recall})}{\mathrm{Precision+Recall}} \tag{3-3}$$

In addition to randomly selecting the training and testing data, this study tested the algorithms with continuous data. In particular, the continuous pattern of acceleration data was used for training models, and the trained model was evaluated with strictly continuous acceleration signals. As continuous acceleration signals reflect actual construction tasks better than randomly selected data, the prediction results are supposed to show more realistic action recognition performance in practice. Postprocessing techniques were applied to benefit from this additional information of continuous data (Gil-Martín et al., 2020). On the basis of our preliminary examination of the results, some errors were frequently observed in the middle of ongoing work for a specific activity, and the misclassified data were relatively short, lasting only for 1 or 2 s. Considering the context of the construction activities, this intermittent class found in the classification results will likely be an error. Thus, if the predicted class of the activity 1) lasts less than the unit length of the sliding window and 2) the class is observed in the middle of other continuously lasting activities, then this intermittent class was regarded as a misclassified class, and the class was modified as adjacent classes. After the postprocessing procedure, the study then calculated how much time was spent on each activity, which can potentially help evaluate the productivity of each worker.

## 3.3 Result

### 3.3.1 Accuracy of the trained models

Window size is a crucial parameter for accelerometer-based activity recognition. This study investigated the window size through pretesting, and the optimal window size was determined as 1.5 s after multiple tests. In the meantime, the study employed a five-fold cross-validation strategy, wherein each trained model was evaluated using the testing dataset to obtain its testing accuracy. Consequently, this approach yielded five testing accuracies corresponding to each of the trained models. The average testing accuracy served as the primary metric for evaluation and is documented in Table 3-1, which shows the performance of classification results for three levels of activities according to 1) classifiers (i.e., traditional machine learning and deep learning algorithms), 2) validation methods (i.e., LOOCV and LOSOCV), and 3) data sampling (i.e., discrete data and continuous data). According to the results from LOOCV, Level 1 classification shows excellent performance with over 90% accuracy, while the deep learning model (i.e., BiLSTM) shows slightly better accuracy than the machine learning model (i.e., Ensemble Bagged Trees). At Level 2, classification results from LOOCV range between 80% and 90%, with the deep learning model demonstrating superior accuracy, particularly for formwork tasks. At Level 3, the deep learning model showed significantly higher classification performance than traditional machine learning, indicating that the use of deep learning algorithms would be recommended to classify complex construction activities. However, the overall accuracy at Level 3 was about 77.0% and 74.9% for formwork and rebar work, respectively, even when using the deep learning model. Evaluating classifiers with continuous data, either through LOOCV or LOSOCV, reveals a significant decrease in overall testing accuracy compared to results obtained from LOOCV with the discrete data. Such a decline suggests considerable variations in data related to collection times and subjects involved.

Table 3-1 Models performance overview

| Work division | | | Formwork | | | Rebar work | | |
|---|---|---|---|---|---|---|---|---|
| Testing data selection | | | LOOCV with discrete data | LOOCV with continuous data | LOSOCV | LOOCV with discrete data | LOOCV with continuous data | LOSOCV |
| Average Testing Accuracy | Machine Learning (Ensemble Bagged Trees) | Level 1 Activity | 96.2% | 95.3% | 93.7% | 95.7% | 96.1% | 93.5% |
| | | Level 2 Activity | 83.8% | 81.2% | 78.5% | 79.5% | 74.6% | 76.6% |
| | | Level 3 Activity | 61.3% | 50.3% | 42.9% | 57.1% | 45.3% | 44.7% |
| | Deep Learning (BiLSTM) | Level 1 Activity | 98.7% | 98.9% | 94.7% | 98.6% | 98.3% | 97.2% |
| | | Level 2 Activity | 90.6% | 81.6% | 77.8% | 86.6% | 79.3% | 77.2% |
| | | Level 3 Activity | 77.1% | 55.7% | 49.0% | 74.9% | 57.7% | 55.6% |

*LOOCV: leave-one-out cross-validation, LOSOCV: leave-one-subject-out cross-validation

Based on the results from LOOCV with discrete data, the confusion matrices of all formwork and rebar work activities, the predicted category, actual category, precision, recall, and F1 score of each activity are presented in Table 3-2 and Table 3-3. As shown in the Level 2 confusion matrix, the majority of incorrect predictions of traveling are reported as coming from rebar installation or form installation. For instance, the Level 2 classification in Table 3-2 shows that 82.6% of the predictions are "Form installation," but such results actually belong to "Traveling." The prediction errors (98.0%) of form installation are misclassifications between form installation and traveling. Given such consequences, the most significant errors are caused by confusion between traveling and rebar or form installation at Level 2 activities. In the Level 3 confusion matrices, the fractions of activity that are misclassified as "Supplement work" are 73.4%, 80.5%, 82.6%, 76.5%, and 83.7% in the negative predictions of "Form placing," "form connecting," "Form preparation," "Transferring materials and tools, and transportation," respectively. As shown in Table 3-3, the same issue is also observed in the activity recognition for rebar work. In this regard, “Supplement work” at Level 3 activities is the most dynamic activity that caused considerable confusion with traveling-related activities and other “Material installation” activities. Such facts might imply that the confusion between “Form” or “Rebar installation” and “Traveling” at Level 2 is mainly due to the confusion between “Supplement work” and traveling-related activities at Level 3.

Table 3-2 Confusion matrix of formwork activity classification

| Level 1 Activity | Predicted category | W | I | Recall (%) |
|---|---|---|---|---|
| True category | W | 25894 | 380 | 98.6 |
|  | I | 1183 | 12759 | 91.5 |
|  | Precision (%) | 95.6 | 97.1 |  |
|  | F1 Score | 1.0 | 0.9 |  |

* W: Work, I: Idling

| Level 2 Activity | Predicted category | W_FI | W_TR | I_SS | Recall (%) |
|---|---|---|---|---|---|
| True category | W_FI | 19889 | 545 | 459 | 95.2 |
|  | W_TR | 4431 | 845 | 89 | 15.8 |
|  | I_SS | 995 | 16 | 12947 | 92.8 |
|  | Precision (%) | 78.6 | 60.1 | 95.9 |  |
|  | F1 Score | 0.9 | 0.3 | 0.9 |  |

* W_FI: Form installation, W_TR: Traveling, I_SS: Stand/sit

| Level 3 Activity | Predicted category | W_FI_SP | W_FI_PL | W_FI_CT | W_FI_PA | W_TR_MT | W_TR_SP | I_SS_ST | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|
| True category | W_FI_SP | 7863 | 2 | 529 | 738 | 36 | 468 | 372 | 78.6 |
|  | W_FI_PL | 643 | 13 | 73 | 100 | 2 | 16 | 42 | 1.5 |
|  | W_FI_CT | 2818 | 1 | 1557 | 374 | 8 | 121 | 180 | 30.8 |
|  | W_FI_PA | 3039 | 2 | 277 | 1005 | 9 | 118 | 234 | 21.5 |
|  | W_TR_MT | 1256 | 0 | 73 | 110 | 82 | 131 | 72 | 4.8 |
|  | W_TR_SP | 2338 | 1 | 149 | 196 | 1 4 | 834 | 109 | 22.9 |
|  | I_SS_ST | 477 | 0 | 58 | 126 | 5 | 25 | 13283 | 95.1 |
|  | Precision (%) | 42.3 | 68.4 | 57.0 | 37.6 | 52.6 | 48.0 | 92.9 |  |
|  | F1 Score | 0.6 | 0.0 | 0.4 | 0.3 | 0.1 | 0.3 | 0.9 |  |

* W_FI_SP: Supplement work, W_FI_PL: Form placing, W_FI_CT: Form connecting, W_FI_PA: Form preparation, W_TR_MT: Transferring materials and tools, W_TR_SP: Transportation, I_SS_ST: Standing/Sitting

Table 3-3 Confusion matrix of rebar work activity classification

| Level 1 Activity | Predicted category | W | I | Recall (%) |
|---|---|---|---|---|
| True category | W | 25001 | 301 | 98.8 |
| | I | 812 | 9622 | 92.2 |
| | Precision (%) | 96.9 | 97.0 | |
| | F1 Score | 1.0 | 1.0 | |

* W: Work, I: Idling

| Level 2 Activity | Predicted category | W_RI | W_TR | I_SS | Recall (%) |
|---|---|---|---|---|---|
| True category | W_RI | 16568 | 1200 | 303 | 91.7 |
| | W_TR | 4911 | 2212 | 104 | 30.6 |
| | I_SS | 682 | 42 | 9714 | 93.1 |
| | Precision (%) | 74.8 | 64.0 | 96.0 | |
| | F1 Score | 0.8 | 0.4 | 1.0 | |

* W_RI: Rebar installation, W_TR: Traveling, I_SS: Stand/sit

| Level 3 Activity | Predicted category | W_RI_SP | W_RI_PL | W_RI_CT | W_RI_PA | W_TR_MT | W_TR_SP | I_SS_ST | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|
| True category | W_RI_SP | 1498 | 73 | 362 | 26 | 3 | 434 | 103 | 59.9 |
| | W_RI_PL | 414 | 240 | 211 | 14 | 1 | 169 | 46 | 21.9 |
| | W_RI_CT | 665 | 86 | 688 | 22 | 0 | 236 | 69 | 39.0. |
| | W_RI_PA | 307 | 39 | 148 | 92 | 0 | 172 | 52 | 11.4 |
| | W_TR_MT | 73 | 5 | 13 | 3 | 21 | 98 | 3 | 9.7 |
| | W_TR_SP | 696 | 43 | 205 | 17 | 2 | 1188 | 56 | 53.8 |
| | I_SS_ST | 135 | 13 | 39 | 9 | 0 | 53 | 3059 | 92.5 |
| | Precision (%) | 39.5 | 48.1 | 41.3 | 50.0 | 77.8 | 50.6 | 90.3 | |
| | F1 Score | 0.5 | 0.3 | 0.4 | 0.2 | 0.2 | 0.5 | 0.9 | |

* W_RI_SP: Supplement work, W_RI_PL: Rebar placing, W_RI_CT: Rebar connecting, W_RI_PA: Rebar preparation, W_TR_MT: Transferring materials and tools, W_TR_SP: Transportation, I_SS_ST: Standing/Sitting

### 3.3.2 Activity time estimation

The activity time estimation was performed to further examine the applicability of the action recognition approach for more detailed activity analysis in the construction field. For the performance measurement, the duration of each activity was first calculated on the basis of the recorded video data. The average duration of formwork was 2.8 h, and the average length of a rebar work was 1.9 h. This study cumulated the prediction results to measure the time spent on each activity category. With the estimated duration of each activity and the ground truth, the performance of activity time estimation was calculated. As shown in Table 3-4, the average estimation accuracy of Level 1 activities is 99.5% and 99.4% for formwork and rebar work, respectively. The trained models can determine the working time of formwork and rebar work with accuracies of 96.6% and 92.0%, respectively. The estimation accuracy of Level 3 activities is 65.2% and 74.4%, respectively. Such results imply the feasibility of monitoring the progress of each activity by utilizing wearable data from the construction environment. In particular, the proposed time estimation method contributes to the precise distinction between effective and ineffective work, and such facts offer an opportunity to implement countermeasures to the activity in question.

Table 3-4 Spending time estimation of Level 1 activity

| Work division | Sample # | | Time (hour) | | Accuracy (%) |
|---|---|---|---|---|---|
| | | | Work | Idling | |
| Formwork | 1 | Ground truth | 1.9 | 0.9 | 99.2 |
| | | Estimation | 1.9 | 0.9 | |
| | 2 | Ground truth | 1.9 | 0.9 | 99.8 |
| | | Estimation | 1.9 | 0.9 | |
| | 3 | Ground truth | 1.9 | 0.9 | 99.6 |
| | | Estimation | 1.9 | 0.9 | |
| | 4 | Ground truth | 1.9 | 0.9 | 99.9 |
| | | Estimation | 1.9 | 0.9 | |
| | 5 | Ground truth | 1.7 | 1.1 | 98.7 |
| | | Estimation | 1.7 | 1.1 | |
| | Average | | | | 99.5 |
| Rebar work | 1 | Ground truth | 1.1 | 0.9 | 99.3 |
| | | Estimation | 1.1 | 0.9 | |
| | 2 | Ground truth | 1.1 | 0.9 | 99.7 |
| | | Estimation | 1.1 | 0.9 | |
| | 3 | Ground truth | 1.1 | 0.8 | 99.3 |
| | | Estimation | 1.1 | 0.9 | |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
|  | 4 | Ground truth | 1.1 | 0.8 | 99.0 |
|  |  | Estimation | 1.1 | 0.8 |  |
|  | 5 | Ground truth | 1.1 | 0.8 | 99.8 |
|  |  | Estimation | 1.1 | 0.8 |  |
|  | Average |  |  |  | 99.4 |

Table 3-5 Spending time estimation of Level 2 activity

| Work division | Sample # |  | Time (hour) |  |  | Accuracy (%) |
|---|---|---|---|---|---|---|
|  |  |  | W_MI* | W_TR | I_SS |  |
| Formwork | 1 | Ground truth | 1.5 | 0.4 | 0.9 | 97.2 |
|  |  | Estimation | 1.5 | 0.4 | 0.9 |  |
|  | 2 | Ground truth | 1.5 | 0.4 | 0.9 | 97.4 |
|  |  | Estimation | 1.5 | 0.4 | 0.9 |  |
|  | 3 | Ground truth | 1.5 | 0.4 | 0.9 | 93.1 |
|  |  | Estimation | 1.5 | 0.3 | 0.9 |  |
|  | 4 | Ground truth | 1.4 | 0.4 | 0.9 | 99.0 |
|  |  | Estimation | 1.4 | 0.4 | 0.9 |  |
|  | 5 | Ground truth | 1.3 | 0.4 | 1.1 | 96.5 |
|  |  | Estimation | 1.4 | 0.3 | 1.1 |  |
|  | Average |  |  |  |  | 96.6 |
| Rebar work | 1 | Ground truth | 0.7 | 0.4 | 0.9 | 96.8 |
|  |  | Estimation | 0.7 | 0.3 | 0.9 |  |
|  | 2 | Ground truth | 0.7 | 0.4 | 0.9 | 95.2 |
|  |  | Estimation | 0.7 | 0.3 | 0.9 |  |
|  | 3 | Ground truth | 0.7 | 0.4 | 0.8 | 96.2 |
|  |  | Estimation | 0.8 | 0.3 | 0.8 |  |
|  | 4 | Ground truth | 0.7 | 0.4 | 0.8 | 89.2 |
|  |  | Estimation | 0.8 | 0.3 | 0.8 |  |
|  | 5 | Ground truth | 0.7 | 0.4 | 0.8 | 82.6 |
|  |  | Estimation | 0.9 | 0.3 | 0.8 |  |
|  | Average |  |  |  |  | 92.0 |

* W_MI: Material (formwork and rebar) installation, W_TR: Traveling, I_SS: Stand/sit

Table 3-6 Spending time estimation of Level 3 activity

| Work division | Sample # | | Time (hour) | | | | | | | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | W_FI_SP* | W_FI_PL | W_FI_CT | W_FI_PA | W_TR_MT | W_TR_SP | I_SS_ST | |
| Form work | 1 | Ground truth | 0.6 | 0.0 | 0.4 | 0.5 | 0.2 | 0.2 | 0.9 | 59.3 |
| | | Estimation | 1.0 | 0.0 | 0.3 | 0.2 | 0.1 | 0.3 | 0.9 | |
| | 2 | Ground truth | 0.6 | 0.0 | 0.4 | 0.5 | 0.2 | 0.2 | 0.9 | 67.0 |
| | | Estimation | 0.9 | 0.0 | 0.4 | 0.2 | 0.1 | 0.3 | 0.9 | |
| | 3 | Ground truth | 0.6 | 0.0 | 0.4 | 0.5 | 0.2 | 0.2 | 0.9 | 61.0 |
| | | Estimation | 0.9 | 0.0 | 0.3 | 0.2 | 0.1 | 0.3 | 0.9 | |
| | 4 | Ground truth | 0.6 | 0.1 | 0.4 | 0.4 | 0.2 | 0.2 | 0.9 | 67.8 |
| | | Estimation | 0.7 | 0.0 | 0.3 | 0.3 | 0.1 | 0.4 | 1.0 | |
| | 5 | Ground truth | 0.0 | 0.5 | 0.0 | 0.2 | 0.5 | 0.1 | 0.3 | 70.4 |
| | | Estimation | 0.0 | 0.7 | 0.1 | 0.4 | 0.2 | 0.1 | 0.3 | |
| | Average | | | | | | | | | 65.2 |
| Rebar work | Sample # | | W_RI_SP** | W_RI_PL | W_RI_CT | W_RI_PA | W_TR_MT | W_TR_SP | I_SS_ST | Accuracy (%) |
| | 1 | Ground truth | 0.2 | 0.1 | 0.2 | 0.2 | 0.0 | 0.4 | 0.9 | 84.6 |
| | | Estimation | 0.2 | 0.0 | 0.2 | 0.2 | 0.0 | 0.4 | 0.9 | |
| | 2 | Ground truth | 0.2 | 0.1 | 0.2 | 0.2 | 0.0 | 0.4 | 0.9 | 91.4 |
| | | Estimation | 0.2 | 0.1 | 0.2 | 0.2 | 0.0 | 0.3 | 0.9 | |
| | 3 | Ground truth | 0.2 | 0.2 | 0.2 | 0.2 | 0.0 | 0.4 | 0.8 | 66.2 |
| | | Estimation | 0.0 | 0.4 | 0.1 | 0.2 | 0.0 | 0.0 | 0.4 | |
| | 4 | Ground truth | 0.2 | 0.2 | 0.2 | 0.2 | 0.0 | 0.4 | 0.8 | 68.2 |
| | | Estimation | 0.0 | 0.3 | 0.1 | 0.3 | 0.0 | 0.0 | 0.4 | |
| | 5 | Ground truth | 0.2 | 0.2 | 0.2 | 0.2 | 0.0 | 0.4 | 0.8 | 61.8 |
| | | Estimation | 0.0 | 0.4 | 0.2 | 0.3 | 0.0 | 0.0 | 0.3 | |
| | Average | | | | | | | | | 74.4 |

* W_FI_SP: Supplement work, W_FI_PL: Form placing, W_FI_CT: Form connecting, W_FI_PA: Form preparation, W_TR_MT: Transferring materials and tools, W_TR_SP: Transportation, I_SS_ST: Standing/Sitting

** W_RI_SP: Supplement work, W_RI_PL: Rebar placing, W_RI_CT: Rebar connecting, W_RI_PA: Rebar preparation, W_TR_MT: Transferring materials and tools, W_TR_SP: Transportation, I_SS_ST: Standing/Sitting

## 3.4 Discussions

### 3.4.1 Feasibility of acceleration-based activity recognition in the construction field

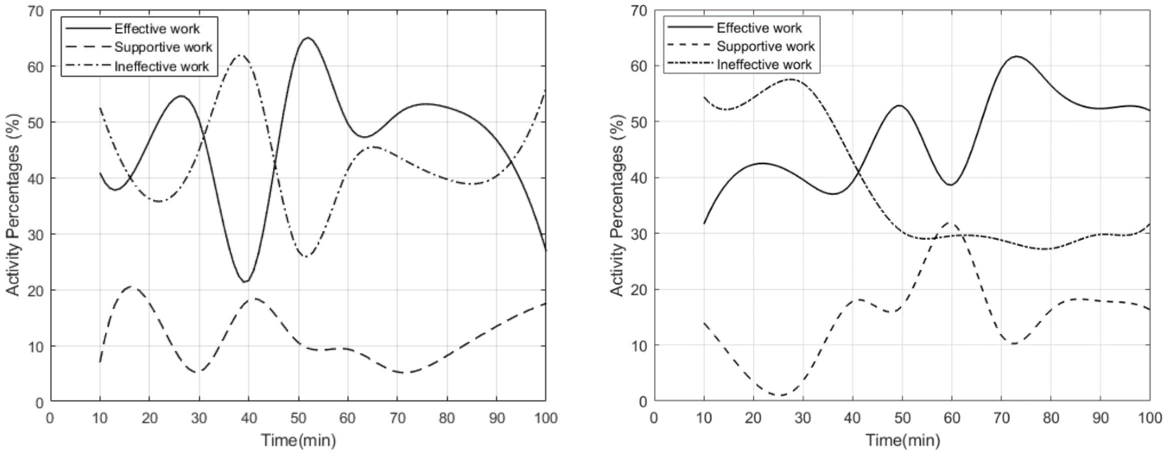
Previous research showed the potential of acceleration-based activity recognition to recognize diverse construction activities. However, the applicability of field activity detection has not been validated in terms of 1) the reliability of activity recognition in field conditions and 2) the defining of construction activities. The activity recognition algorithms in previous studies have been tested with discrete or independent data that ignore the noise and sequence characteristics of continuous acceleration signals collected from construction job sites. Construction activities in previous research are categorized on the basis of single standards, such as the nature of movement or contribution of tasks. Therefore, the derived classification results have limitations in providing information for measuring the efficiency of construction workers or for finding low productivity areas in the construction field concerned. We propose a new taxonomy to address these issues with consideration of movement and work context and subsequently validate it by using extensive field data.

The understanding of the exclusive characteristics of different human activities is challenging due to the complex nature of human activities, which can induce classification confusion. Therefore, defining activities with a clear and comprehensive understanding of their nature is necessary for developing useful activity taxonomy (Bulling et al., 2014). Previous attempts in activity definition have primarily oriented toward a single principle (e.g., nature of the movement or contribution of work), and classifications of construction activities based on such principle have been validated in many previous studies. (Akhavian & Behzadan, 2016; Joshua & Varghese, 2014; Ryu et al., 2019; Weiss et al., 2016). Although movement-based activity taxonomy has a high classification accuracy, it still has several limitations when dealing with practical problems. First, depending on the context, similar movements can be delivered from different activities. In this case, the classification algorithms will perform poorly, especially when the activities being classified have largely similar movement characteristics. Second, a movement-based activity taxonomy (e.g., lifting, sitting, and walking) cannot deliver sufficient information to solve practical problems, such as identifying low-productivity operations in the field.

To overcome these issues, researchers in several studies have introduced a context-based activity taxonomy. The taxonomy categorizes construction activities based on their contributions to the project (Forde & Buchholz, 2004; Hallowell & Gambatese, 2009; Joshua & Varghese, 2014), allowing for the evaluation of productivity in a rough manner. However, most construction activities consist of diverse tasks (e.g., the effective work of an ironworker includes fetching, adjusting, and tying rebar). Previous context-based activity taxonomies are insufficient to reveal the root causes of low productivity due to the lack of detailed information about ongoing activities. In an attempt to solve such problems, this study considered movement- and context-based taxonomy when defining an activity. Theoretically, acceleration signals collected from the dominant hand are regarded as an integrated response of whole-body movements and hand movements (Ryu et al., 2019). Therefore, a different combination of body and hand movement is an intuitive standard for identifying activities that share a distinct acceleration response. However, activities that have similar movements (e.g., lifting material from the ground, squatting, and standing up) are difficult to identify accurately in accordance with the movement-based system. The context standard was introduced to enrich the textural information of the activity and to extend the classification categories. In this regard, the capability of activity recognition for identifying low productivity issues is enhanced.

The construction activities are formatted as a three-level taxonomy with a hierarchical structure (Table 2-1), which allows classifying specific activities by zooming in or out the action level and identifying the optimal classification level by trading off between performance (i.e., accuracy) and outcomes (i.e., information extracted from the results) (Blanke & Schiele, 2010; Krishnan et al., 2013). On the basis of the result shown in Table 3-1, the neural network algorithms can train more powerful classifiers. The classification accuracy at Level 1 (i.e., “Idling” and “Work”) shows over 90% accuracy because “Idling” involves mostly no movement of hands, which can be easily distinguished from “Work,” which involves significant arm and body movements and has substantial changes in acceleration signals. At Level 2, we further divide "Work" into two subcategories, 1) traveling and 2) installing tasks, considering that they have different work contexts (e.g., traveling is a supportive activity, and installing material is a value-added task) and body movements (e.g., "Traveling" involves abundant body movements and few cyclic movements from hands, and "Installing" involves abundant hand movements and few body movements). The classification accuracy at Level 2 is over 80%, and the algorithm can differentiate between horizontal whole-body movements (e.g., "Traveling") and hand-dominant activities (e.g., "Material installation"). In accordance with the confusion matrix at this level (Table 3-2 and Table 3-3), the most significant errors result from the confusion between "Traveling" and "Material installation" because "Material installation" frequently involves a temporal allocation (e.g., moving 1–2 m to pick up materials), which has a large similarity with "Traveling" (e.g., moving to another work zone). The accuracy of Level 3 activity classification is lower than that of Level 1 and Level 2, showing 50%–60% accuracy because more detailed work contexts were contained. The classification results show that significant confusion within the offspring categories of Level 2 activity, "Material installation," occurs. This finding may indicate that the proposed algorithm cannot recognize the considerable interclass variability in Level 3 activities due to the similar nature of body and hand movements for these activities. As the types of activities at Level 3 were more frequently changed during the operation, the acceleration signals may include the noise data from transition patterns between activities. However, in terms of measuring spending time for Level 3 activities, the accuracy increased up to approximately 75% (Table 3-6), showing the potential for being used to understand productivity issues during construction operations.

The classification results at Level 2 are accurate, allowing the identification of productivity issues by providing meaningful information, such as the time expenditure of workers. For instance, two continuous patterns of acceleration data were sampled from two form workers who were at the same site and worked simultaneously. The activity percentage values were calculated on the basis of the spending time estimation method in section 3.3.2, and the percentages were plotted in a time series domain, as shown in Figure 3-4. In particular, the activity percentages of the two form workers were calculated on the basis of 10 min. The productivity of form worker No.2 was higher in the selected 100 min because his effective work rate remained relatively high without any huge drop by comparing Figure 3-4 (a) and Figure 3-4 (b). The cause of the low productivity issues can be exposed. The effective work rate of worker No.1, as shown in Figure 3-4 (a), dropped from 30 min to 40 min, while the ineffective rate increased significantly in the same period. This finding indicates that the increasing proportion of ineffective work is the cause of the low productivity issue in the selected period. The root cause of the low productivity issue of form worker No.2 from 50 min to 60 min can be recognized as the increasing percentage of supportive work by using the same method. Considering the ineffective work is not dominant, and the effective work rate remains at 40%, the worker was on short travel between two installation trades.



a. Form worker No.1 b. Form worker No.2

Figure 3-4 Time series line plot illustrating activity percentage every 10 min

### 3.4.2 Remaining challenges to enhance the classification performance

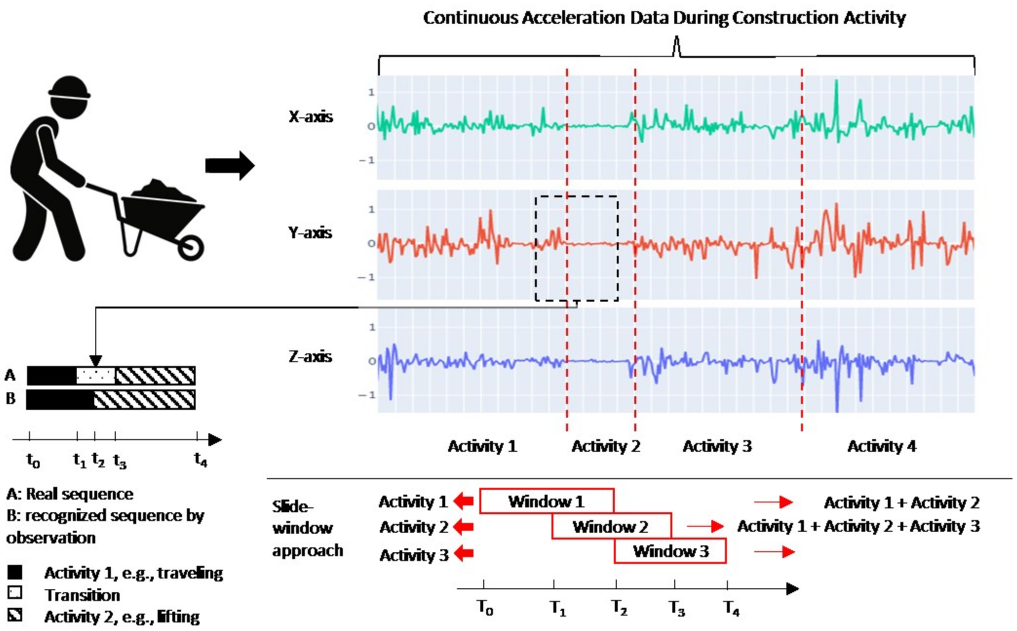
Although the classification result at Level 2 activity can distinguish low productivity issues, it is insufficient to expose the root cause. In this regard, the Level 3 activity is necessary to find the cause of the delay. However, the current performance of Level 3 activity classification does not satisfy the demand in the construction field because recognizing a sequence of activities from an uncontrolled environment (i.e., the construction field) is challenging. In addition to the human variability, several remaining challenges exist, and they are 1) difficulty in handling the transition effect between activities, 2) inaccurate segmentation of time-series movement data, and 3) information loss during the machine learning process. The first challenge deals with the transition moment in continuous human activities (Minnen et al., 2006). In Figure 3-5 (a), sequence A refers to a real activity stream, which indicates that a transition pattern (i.e., pattern from $t_1$ to $t_3$) shall exist between two explicit activities (e.g., traveling and lifting), considering that human activity changes gradually. However, such transition has been disregarded in this study because 1) the duration of the transition activities is relatively short compared with other activities that are explicitly defined in the taxonomy in Table 2-2 (Lara & Labrador, 2012); 2) the temporal boundaries of transitions are difficult to determine by human observation because the transition activity and its neighboring activities share similar movements as recorded in videos. A sample of a labeled sequence (i.e., sequence B) can be found in Figure 5 (a), which shows that activity 1 lasts from $t_1$ to $t_2$, and the following activity (i.e., activity 2) lasts from $t_2$ to $t_4$. A comparison between the real sequence (i.e., sequence A) and the recognized sequence (i.e., sequence B) shows that the two transition patterns (i.e., activity from $t_1$ to $t_2$ and activity from $t_2$ to $t_3$) are mistakenly recognized as activity no.1 and activity no.2, respectively. Considering the transition effect is widespread in the continuous activity patterns, the massive mislabeling of the activity category induces significant errors when training the dataset and the ground truth. Thus, the misclassification rate is considerably high, and the classification system is unacceptable for field productivity evaluation.

One of the alternatives is to regard “Transition” as an extra activity to address this issue (Zhang et al., 2010). In previous research, Rednic et al. (2013) used a transition filter to improve classification accuracy and stability. On the basis of the assumption that more recent posture has a higher correlation with the actual posture, the weighted-voting methods can filter out unreasonable postural vibrates located in the high-frequency domain. The filtering process is validated as useful for increasing the certainty of the transition boundaries. However, the improvement in accuracy is limited. Rather than setting clear-cut boundaries, some researchers (Abonyi et al., 2005) introduced the idea of fuzzy clustering (i.e., data points can belong to more than one cluster) that helps to determine the fuzzy boundaries of time-series data (e.g., the continuous acceleration data). Fuzzy segmentation (i.e., setting fuzzy boundaries for the activity pattern) is then adopted in the activity recognition to overcome the transition effect (Zhang et al., 2014). The researchers defined the fuzzy boundaries with Gaussian membership and a time variable and translated the segmenting issue into an optimizing problem. The bias caused by the transition effect can be restricted by solving the optimization problem. In future research, we will apply the proposed approaches and test the feasibility of reducing transition effect in continuous field data.

In the classification of human activities, continuous sensor data are segmented into sequences for the feature extraction process. However, the setting of data windows of activities without introducing any classification errors is still a challenging task (Bao & Intille, 2004). A sliding window technique for data segmentation was primarily applied, investigated, and validated in previous research (Bulling et al., 2014). Similar to previous studies, we used a sliding window technique with a fixed window size. As shown in Figure 3-5 (b), the acceleration data collected during construction activities (i.e., activities from $T_0$ to $T_5$) are segmented into three windows (i.e., independent activity pattern). Specifically, window No.1 lasts from $T_0$ to $T_2$, window No.2 lasts from $T_1$ to $T_3$, and window No.3 lasts from $T_2$ to $T_4$. The duration of the windows (i.e., $T_0$ to $T_2$, $T_1$ to $T_3$, $T_2$ to $T_4$) is constant, and the overlapping between two consequent windows is set to 50%. However, the use of the fixed-size sliding window can induce considerable misclassification due to two causes of errors (Gu et al., 2009). The duration of the different activity categories is diverse due to the different natures of human movement. Spending time on the same type of activity can vibrate during work. In these regards, a fixed-size window cannot purely and fully include a single type of activity, leading to extreme errors when preparing training data and testing data. Therefore, enhancing classification performance by window size optimization is difficult (Huynh & Schiele, 2005). Previous research demonstrated that the algorithms could perform better if the features and length of windows were considered as separate activity categories.

The multiclass problem is another observed issue related to the sliding window approach (Yao et al., 2018). As shown in Figure 3-5 (b), multiple categories of activity can be found in the same window (e.g., window No.1 consists of activity no.1 and activity no.2; and window No.2 includes activity no.1, activity no.2, and activity no.3). However, following the majority voting principle, a single activity label should be assigned to each data window, which can bring about a significant loss of activity information and result in considerable misclassification. The ground truth of the activity may be disturbed because the true label is different from the label selected for the window. For instance, the data for activity no.2 was labeled as activity no.1 in the segmenting process in window No.1 in Figure 3-5 (b). Therefore, the data of activity no.1 were accidentally polluted by the activity 2 data, resulting in the misleading of the algorithms. Laguna et al. (2011) proposed a dynamic segmenting approach to address these limitations. In this approach, the starting and end times of the window and the window length are concluded as core parameters to determine the windows dynamically. Therefore, changes in activities are integrated into formulas as a significant variable for indicating the beginning and ending points of the window. The results show that the dynamic window approach effectively reduces classification confusion. Yao et al. (2018) proposed a dense labeling scheme that labels each data point rather than labeling the data segment. Each data point can be regarded as a "window" that includes only one datum. The data point is assigned a unique label that any vote-based filtering will not adjust. Therefore, the problems of information loss and label confusion caused by the sliding window method can be overcome.

The last issue of the current model is that the sequential characteristic of continuous construction activity is still ignored. In a sequential activity for construction (i.e., activities that occur in a certain order), an activity can affect the action that occurs after it. For instance, if the prior activity is "Sitting," then the subsequent behavior cannot be "Walking" or "Running" because the activity "Standing up" cannot be avoided between "Sitting" and "Walking." A transition from "Walking" to "Standing up" is also impossible based on the context. In this study, such unreasonable sequences are frequently observed from the classification model, resulting in significant errors. To overcome this issue, Panahandeh et al. (2013) introduced the continuous Hidden Markov Model (HMM) to analyze gait phase and joint activity via IMU measurements. Five individual activities, namely, going upstairs, going downstairs, running, standing, and walking, are discussed in the study. The HMM model integrates the activity influence through two objects: 1) a discrete chain of activities, which reflects the order and relationship between activities, and 2) probability density functions of the future variables, which add the influence on the classification algorithms. The final classification accuracy of this probabilistic activity ranges from 90% to 99%, indicating a great potential for solving the continuous human activity classification problem. Future research can test the continuous HMM with the field-collected data to reduce any unreasonable sequences existing in the classification results.


Continuous Acceleration Data During Construction Activity
X-axis
Y-axis
Z-axis
1
0
−1
Activity 1
Activity 2
Activity 3
Activity 4
A
B
$t_0$ $t_1$ $t_2$ $t_3$ $t_4$
A: Real sequence
B: recognized sequence by observation
■ Activity 1, e.g., traveling
□ Transition
▨ Activity 2, e.g., lifting
Slide-window approach
Activity 1
Activity 2
Activity 3
Window 1
Window 2
Window 3
Activity 1 + Activity 2
Activity 1 + Activity 2 + Activity 3
$T_0$ $T_1$ $T_2$ $T_3$ $T_4$

a. transition effect b. sliding window approach

Figure 3-5 Illustration of errors induced by transition effect and segment method

## 3.5 Conclusions

This study investigated the validity of action recognition algorithms with a newly proposed comprehensive and universally applicable work taxonomy that was designed considering movement and construction contexts. In particular, the performance of the proposed approach was studied by using acceleration data collected in a construction site during unstructured ongoing concrete work. Acceleration signals during formwork and rebar work were labeled with activities defined at three hierarchical levels based on the proposed activity taxonomy and used for testing traditional machine learning- and deep learning-based action recognition algorithms. The testing results show that the classification performance for Level 1 activities for formwork and rebar work is relatively reliable, with accuracy higher than 95%, and the prediction accuracy ranges from 74.6% to 83.8% for Level 2 activity classification. The classification accuracies for Level 3 activities vary from 45.3% to 61.3%.

The classification results for activities at Level 1 and Level 2 demonstrate that 1) the proposed taxonomy can convey comprehensive activity information (i.e., activity context information and movement information) and reduce confusion among the categories in the same level, and 2) the performance of acceleration-based activity recognition algorithm is acceptable when dealing with noisy data (i.e., long-term and continuous data collected directly from the construction site). However, the rather low accuracy for activities at Level 3 may indicate the limitation of the use of acceleration signals for micro-level activity analysis. This study evaluated the spending time estimation of long-term continuous signals collected from the field, which reported high accuracies in measuring the activity duration of Level 1 and Level 2 activity. On the basis of the duration data, the time spent ratio of each activity can be evaluated through the timeline. Therefore, evaluating the work efficiency is possible by comparing it with the benchmark. The root cause of the low-efficiency problem can be exposed by analyzing the time spent ratio, which will help optimize the construction trade to improve productivity.

Measuring workers' activities can provide quantitative evidence for identifying productivity issues from the perspective of individual workers. Acceleration-based action recognition is regarded as a useful means for automated activity analysis, but it suffers from a nonstandardized definition of activities and a lack of validity in a practical setting. This study may provide a solid foundation for automated activity analysis by proposing a practical approach to defining and analyzing construction activities using acceleration data. The comprehensive validation of action recognition algorithms using unstructured field data in this study can convince practitioners about the reliability of acceleration-based action recognition for Level 1 and Level 2 activities in practice.

# CHAPTER 4 SENSOR FUSION-BASED CONSTRUCTION WORKER ACTIVITY RECOGNITION

## 4.1 Background

Activity recognition utilizing sensor data has become increasingly prevalent in the construction industry (Sherafat et al., 2020). This rise can be attributed to advancements in sensor technology, which have analyzed complex construction activities more easily and accurately (Zhang et al., 2017). Despite the promise of sensor technology, there are inherent limitations when it comes to activity recognition in the dynamically changing environment of a construction site. For example, the effectiveness of surveillance cameras can be compromised by physical obstructions, leading to data missing issue These challenges are primarily due to the limitations of individual sensors in adapting to the diverse demands of a construction environment.

One innovative solution to overcome these limitations is the multisensor fusion approach. This transformative method integrates inputs from various sensors to form a unified model, enhancing precision and reliability. By employing multisensor fusion, the strengths of each sensor type are harnessed, effectively countering the deficiencies observed in single-sensor systems (Ayed et al., 2015; Khaleghi et al., 2013; Kokar et al., 2004). Sensor fusion has gained prominence in the construction industry (Rao et al., 2022), where sensor-based approaches are increasingly used to analyze construction workers' operations timely through advancements in sensor technology (Akhavian & Behzadan, 2016; Luis Sanhudo et al., 2021). Despite the sophistication of sensor technologies, the inherent limitations of specific sensors restrict their capability to collect accurate and reliable data, especially in dynamic environments such as construction sites, thereby compromising the integrity of data analysis. For instance, the efficacy of surveillance cameras can be compromised by obstructions, leading to data loss (Bohn & Teizer, 2010). The multisensor fusion approach has been proposed to mitigate such constraints, integrating inputs from diverse sensors into a cohesive model. Employing the sensor fusion strategy enhances the precision and reliability of the sensor network's inferences. This approach leverages the strengths of each sensor, compensating for the limitations inherent in single-sensor systems (Hall & Llinas, 1997; Khaleghi et al., 2013).

However, the effectiveness of sensor fusion networks in construction for worker activity recognition remains unexplored. Furthermore, the potential for accelerometer and camera data to complement each other and thereby enhance sensor fusion effectiveness is unknown. To address these issues, this study introduces a sensor fusion framework specifically designed for construction sites. This framework integrates acceleration and vision sensors into a cohesive network. Based on the Dempster-Shafer theory of evidence, it incorporates weight modifications to address the issue of uneven credibility in sensor results. The efficacy of this novel framework has been confirmed through laboratory tests, demonstrating its potential to improve activity recognition in construction environments significantly.

## 4.2 Sensor Fusion Definition

Data fusion is a collection of techniques combining data from numerous sources, aiming to enhance the performance of specific tasks, such as accuracy, stability, and efficiency, compared to using a single data source (Malhotra, 1995; White, 1987). In the 1960s, data fusion terminology was first found in mathematical models for processing data (Esteban et al., 2005). In the 1970s, the data fusion technique gained attention from the US Department of Defense (DoD), which established the Data Fusion Sub-Panel of the Joint Directors of Laboratories (JDL) in order to unify terminology, build principles, address issues, and develop software

(Hall & Llinas, 1997). Therefore, data fusion was first implemented for military purposes, such as target identification and tracking, threat recognition, and battlefield monitoring (Abidi & Gonzalez, 1992). With the technology transferred from the military domain, data fusion technology has also flourished in the civilian domains, such as traffic monitoring (Liu et al., 2013), fault detection (Sarkar et al., 2014), navigation (Qu et al., 2021), autonomous driving (Yeong et al., 2021), and Human Activity Recognition (Qiu et al., 2022). In recent years, the rapid development of sensor technologies has allowed data acquisition to become efficient, accurate, and automated (Sathe et al., 2013; Tubaishat & Madria, 2003). In this regard, the terms "sensor fusion" and "data fusion" are mentioned as being equal when using sensor data as the data source (Elmenreich, 2002). Meanwhile, an extended terminology, "multi-sensor data fusion," was proposed as a technology that concerns how to combine data from multiple sensors, which is the recent definition of sensor-based data fusion (i.e., sensor fusion) (Hall & McMullen, 2004; Waltz & Llinas, 1990). Therefore, this research uses the terms data fusion, sensor fusion, and multisensor data fusion synonymously.

## 4.3 Sensor Fusion Architectures

Various methods of multisensor fusion have been proposed and validated across a broad spectrum of domains (Ayed et al., 2015; Kong et al., 2020; Qiu et al., 2022; Shao et al., 2021). The previous applications suggested that each fusion method has different feasibility toward different tasks. The fusion performance is also prone to be affected by data sources. Therefore, choosing an appropriate fusion approach plays an essential role in ensuring the effectiveness of applying multi-sensory fusion. Previous researchers have established various multi-sensor fusion architectures based on the characteristics and applicability of fusion methods (Ayed et al., 2015), resulting in an efficient pipeline to select a suitable fusion method. One of the earliest structures is the well-known JDL model, which was proposed by White (1987) with the help of the DoD. The conceptual architecture, as depicted in Figure 4-1, comprises three principal components: the data source, the data fusion module, and the Human-Computer Interface. The data fusion module also includes data preprocessing, four steps of data processing (i.e., object refinement, situation refinement, threat refinement, and process refinement), and a database management system (De Boer, 2002).

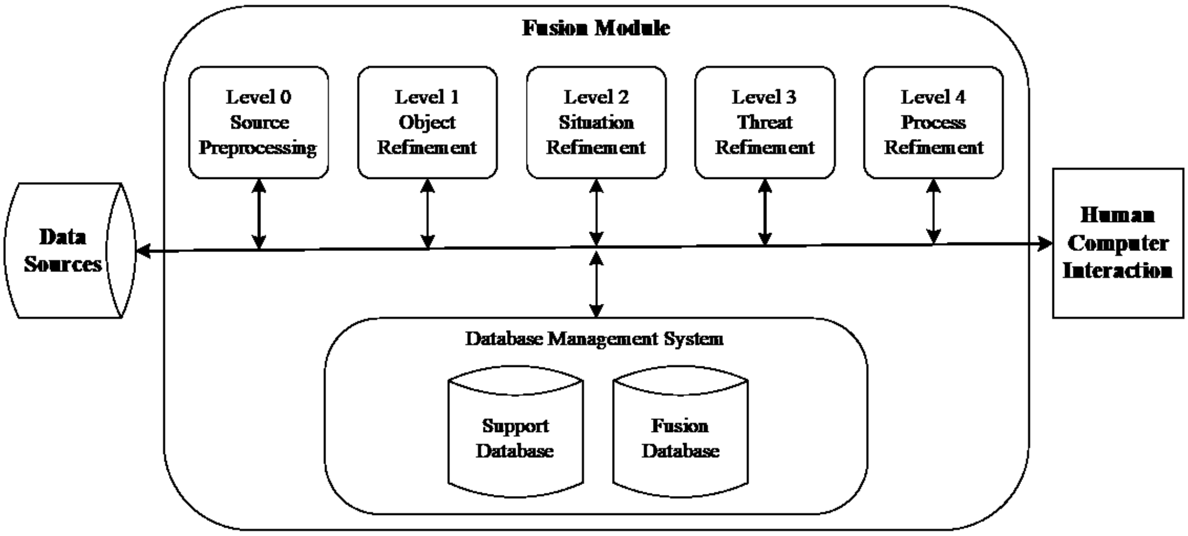

Fusion Module
Level 0 Source Preprocessing
Level 1 Object Refinement
Level 2 Situation Refinement
Level 3 Threat Refinement
Level 4 Process Refinement
Data Sources
Human Computer Interaction
Database Management System
Support Database
Fusion Database

Figure 4-1 Diagram of JDL fusion architecture

The JDL model initially focuses on military applications, such as ocean surveillance, target identification, and improvement of battlefield situational awareness. Due to the existing gap between military and civilian purposes, additional revisions have been made to utilize the civilian domain system (Blasch & Plano, 2002; Llinas et al., 2004; Steinberg & Bowman, 2017; Steinberg, 1999). For instance, Bowman and Morefield (1980) investigated the duality between resource management and data fusion, specifically in Level 2 of the JDL architecture, and then designed a two-level architecture for the resource management demand. Despite its widespread use, the JDL model focuses more on the input and output data rather than processing them. Alternatively, Luo and Kay (1988) proposed a hierarchical multi-sensory fusion architecture that includes the data processing level, known as signal-level, pixel-level, feature-level, and symbol-level, from low to high level (Figure 4-2). The scale on the right side of the diagram interprets the data processing level, and the left side introduces the fusion process. As shown in the figure, $n$ sensors (i.e., $S_1$, $S_2$, ... $S_n$) are installed in the environment, and their data are denoted as $x_1$, $x_2$, …, $x_n$, respectively. The fusion node 1 in the left hand represents a lower level (e.g., signal-level) fusion process, which integrates the raw data from $S_1$ and $S_2$ into a new output, $x_{1,2}$. The raw data $x_3$ could also be fused with $x_{1,2}$ in the subsequent fusion node, resulting in a new representation $x_{1,2,3}$ at a higher level. Likewise, each sensor data can be integrated into the system at the demand level. Besides the fusion level, multiple external parameters could affect the fusion results in practice. Luo and Kay (1988) introduced them as an integration system in their fusion architecture comprising three main functions: 1) sensor selection, 2) world model, and 3) data transformation. The sensor selection refers to the optimal configuration of the sensor network, including but not limited to sensor types, numbers, and deployment. The authors recommended two approaches toward the most appropriate sensor network construction: preselection and real-time selection. The preselection method suggests setting up an initial sensor network first and then adding or moving the sensor nodes by considering the available sensor elements and actual geometric environments (Beni et al., 1983). In contrast to preselection, the real-time selection approach allows subsequent sensors to be arranged with a minimum initialization expense (e.g., one sensor) (Hutchinson et al., 1988). The world model is used to store the data associated with the possible operating environments. The stored data includes both the priori data and the newly collected data. Sensory data processing is more convenient with the given information and environment, particularly in high-level fusion applications (e.g., symbol-level and feature-level) (Luo et al., 2002). Meanwhile, the data transformation function is designed to unify data modalities so that the data acquired from different sources can be fused at the designed level. The signal-level fusion is also known as low-level fusion, data-level, or raw data-level fusion in other literature (Blasch et al., 2010; Kaempchen et al., 2005; Kam et al., 1997), which attempts to integrate the sensory data directly and propagate the information to fusion modules as the input (Figure 4-3 a). Previous studies suggested that homogeneous groups of data sources are preferred as an input of the signal-level fusion method. For example, the pixel-level input in Luo and Kay (1988) structure targets image data operation, such as segmentation tasks. Subsequently, features derived from signals or pixels function as fused objects at the feature-level. This process is designed to enhance the precision of feature measurement.

Feature-level fusion, also identified as medium- and characteristic-level fusion in the context of information representation (Figure 4-3 b), coexists with symbol-level fusion. The latter focuses on integrating outcomes derived from numerous classifiers into a singular resolution. Hence, symbol-level fusion is alternatively referred to as decision-level fusion (Kirstein, 2013; Xu et al., 2016). The framework of decision-level fusion is illustrated in Figure 4-3 c.

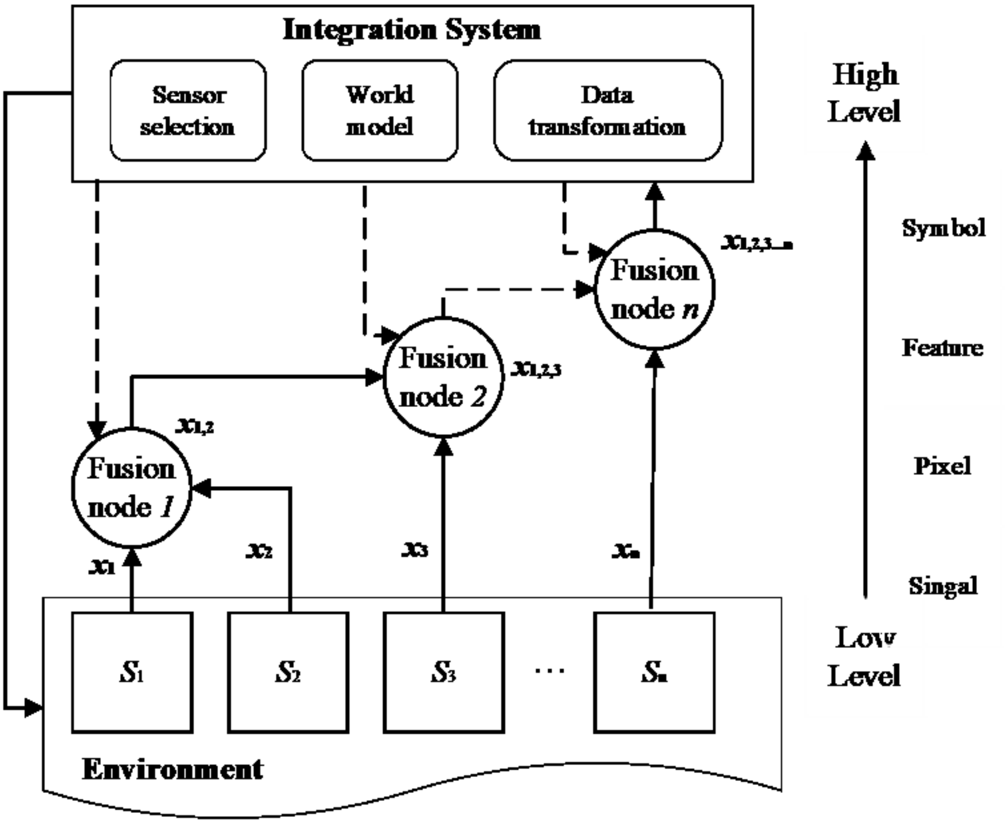
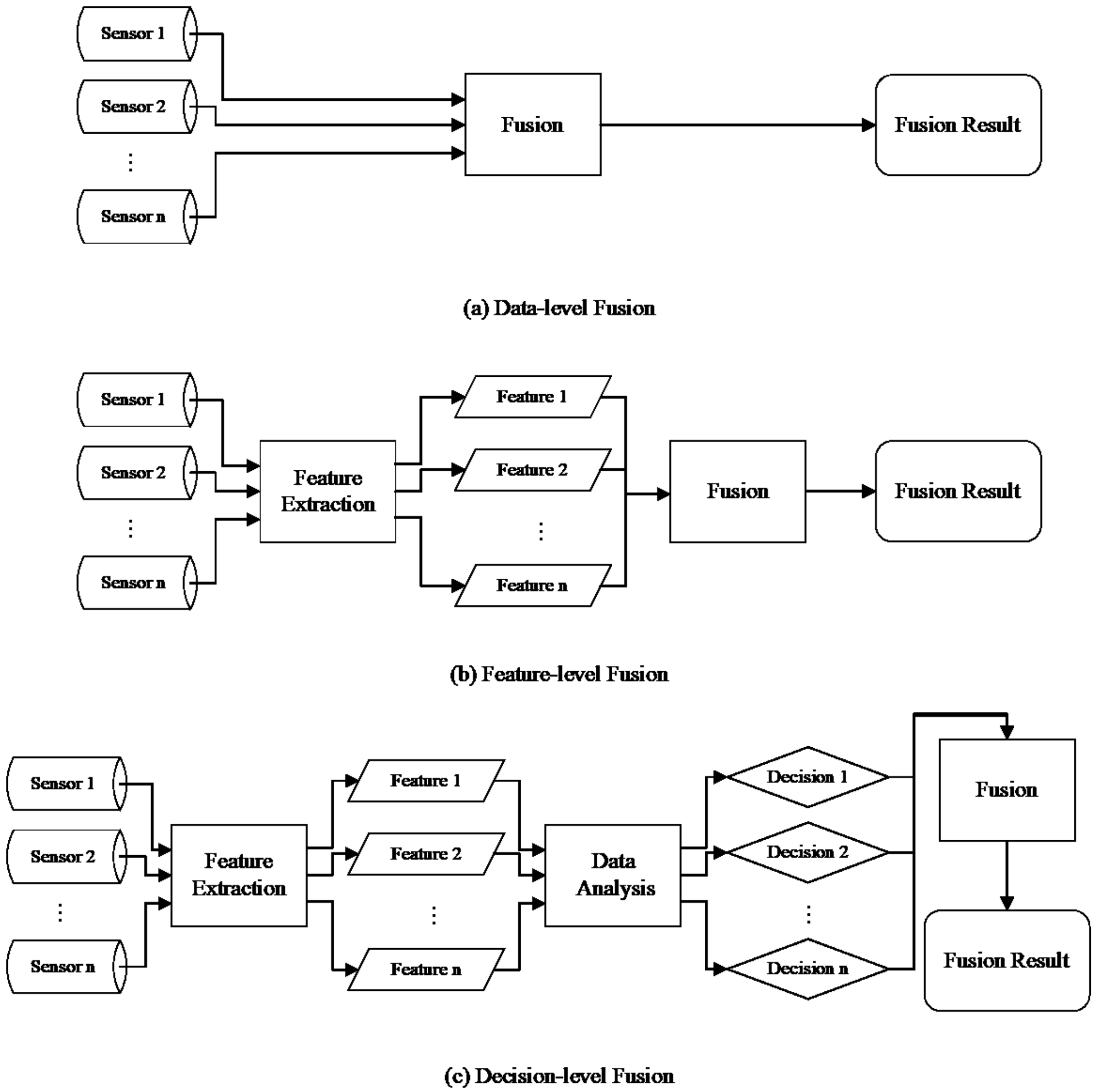


Figure 4-2 Diagram of Luo and Kay (1988)Architecture



(a) Data-level Fusion

(b) Feature-level Fusion

(c) Decision-level Fusion

Figure 4-3 Three multisensor fusion schemes in terms of data processing level

Another general conceptualization was proposed by Dasarathy (1994), who defined the fusion process as a data flow classified by input and output properties. Therefore, this architecture is also named as I/O-Based characterization, which includes five classes: Data In-Data Out (DAI-DAO) fusion, Data In-Feature Out (DAI-FEO) fusion, Feature In-Feature Out (FEI-FEO) fusion, Feature In-Decision Out (FEI-DEO) fusion and Decision In-Decision Out (DEI-DEO) fusion. Other researchers also proposed novel work related to fusion architecture. For instance, Laboratory Analysis Architecture Systems (LAAS) were introduced in 1998. The system was designed to support the implementation of mobile robots in real-time. Another comprehensive work was conducted by Kokar et al. (2004), who used the category theory to establish a general formalization of the fusion system. The proposed system is supposed to cover all kinds of fusion methods, including data-, feature-, and decision-level fusion. Since the 1980s, various architectures have been developed for multisensory fusion, and comparison work among the architectures has been conducted in previous work (Ayed et al., 2015). Upon reviewing the architectures mentioned, the system developed by Luo and Kay (1988) is found to be more frequently utilized in multisensory fusion research. Their taxonomy emphasizes the sensors and categorizes methods according to the processing level of sensor data. This study employs the widely-adopted taxonomy developed based on Luo and Kay (1988) structure, which is a three-level hierarchy of fusion systems: 1) data-level, 2) feature-level, and 3) decision-level fusion.

## 4.4 Sensor Fusion Techniques Selection

Data-level fusion relies on each sensor to capture raw data, which are then integrated for further analysis. In order to successfully fuse data at this level, the raw data are required to be commensurate and associated properly before fusing. As a result, the computational cost is higher than the feature- and decision-level fusion methods (Kulkarni & Rege, 2020). Feature-level fusion architecture is concerned with extracting and fusing feature vectors from each sensor's observations. The feature vectors are synthesized into a single, comprehensive feature vector, which is then processed using techniques such as neural networks and offers a comprehensive output based on the fused feature vectors from all sensors (Hall & Llinas, 1997). Using feature engineering approaches, such as a dimensionality reduction method, the complexity and computational cost of the feature-level approach could be lower than the data-level fusion. In addition, this sensitivity to data compatibility is lower than the data-level method. However, this approach may suffer from sacrificing data during the feature selection or filtering process, resulting in potential data loss issues (Badrinath & Gupta, 2009; Vakil et al., 2021). Decision-level fusion architecture is characterized by an individualized approach whereby each sensor generates a preliminary result based on its data. The final result is obtained by integrating the preliminary results of all sensors using techniques such as classical inference, Bayesian inference, or Dempster-Shafer's method (Hall & Llinas, 1997). It is worth noting that this approach offers several advantages, such as computational efficiency and system diversity. However, it also poses some challenges, including the risk of unreliable decision-making and a lack of data detail. The advantages and disadvantages of data-, feature-, and decision-level fusion are summarized in Table 4-1.

Table 4-1 Comparison of data-, feature-, and decision-level fusion

| Fusion type | Advantages | Disadvantages |
|---|---|---|
| Data-level fusion | Comprehensive data preservation | High computational power and data compatibility |
| Feature-level fusion | Dimensionality reduction and incompatible data tolerance | Critical feature choice and potential data loss |
| Decision-level fusion | Computational efficiency and high system diversity | Unreliable decision risk and lack of data detail |

Among these three fusion approaches, decision-level fusion exhibits a significant comparative advantage for recognizing construction workers' activities, as it remains the least susceptible to data loss (Tzirakis et al., 2019). Compared to early-stage fusion approaches like data-level and feature-level fusion, decision-level fusion offers a distinct advantage by being less sensitive to data incompleteness. It leverages local decisions from individual models as inputs, ensuring system functionality even when modalities encounter issues. This method eliminates the need for complex data completeness assurance systems, thus reducing additional effort. Decision-level fusion allows action recognition algorithms to classify activities even when a data source is missing, such as in cases of occluded images. Furthermore, it requires less precise data alignment than lower-level fusion methods, easing the complexity of data preprocessing. The overall complexity of decision-level fusion is also lower, requiring fewer interconnections between modalities. Therefore, using a decision-level fusion approach, particularly in integrating two heterogeneous data sources, is more straightforward than fusing at an early level of data processing (C. Chen et al., 2017). This fusion approach also facilitates easy updating of data and models, as modifications can be easily made independently without affecting the existing sensor network. From this aspect, the flexibility of the system constructed based on a decision-level fusion approach is higher than other approaches (Gunes & Piccardi, 2005). Considering the applicability, feasibility, and flexibility, the current study implements sensor fusion using the decision-level fusion approach. Traditional decision-level fusion methods employ mathematical theories to integrate the information from multimodal data. Some of the widely used theories adopted in decision-level fusion are probabilistic theory (e.g., Bayesian theory), evidential belief theory (e.g., Dempster-Shafer theory), and rough set-based fusion (Castanedo, 2013; Hall & Llinas, 1997). The current study denotes this branch of decision-level techniques as a theory-based approach. Also, considering that machine learning is commonly used for data analytics for multimodality data, this approach is commonly called a learning-based method for decision-level fusion (Meng et al., 2020). Details on each mathematical theory for decision-level fusion are described below.

### 4.4.1 Majority voting and weighted voting

The majority voting method integrates the output from multiple predictions, estimates, or classifications. The voting function allows the most often occurred result as a final result in the fusion process, which could be illustrated in Equation (4-1). The $X_{\mathrm{i}}$ is the $i^{\mathrm{th}}$ sensor, and $X_{\mathrm{j}}$

represents the $j^{\text{th}}$ hypothesis result from a $1 \times k$ decision vector. Take a multiclass classification problem as an example. $X_j$ refers to the prediction of the $j^{\text{th}}$ label. Then $X_{i,j}$, for $j$=1 to $k$, is a row vector that contains the prediction from one set of observations. By using the *argmax* function Equation (4-2), the fused decision could be obtained when the likelihood reaches the largest value (Sinha et al., 2008).

$$Y_j = \sum_{i=1}^{n} X_{i,j} \ \ \forall j = 1{:}\,k \tag{4-1}$$

$$d = \operatorname{argmax} Y_j \tag{4-2}$$

The majority voting method would be valid only when the decision from each sensor shares the same reliability, uncertainty, and contribution. In other words, all the sensor outputs should have equal priori probabilities. In order to overcome such limitations, the concept of weight is introduced into the voting system (Benediktsson & Kanellopoulos, 1999). The formula is thereby updated as Equation (4-3), where $w_i$ quantifies the weight toward sensor $X_i$, and the sum of weights (i.e., $w_i$, for $i$ = 1 to $n$) is supposed to be one.

$$Y_j = \sum_{i=1}^{n} w_i X_{i,j} \ \ \forall j = 1{:}\,k \tag{4-3}$$

The voting principle is commonly used in decision-level fusion in the multisensory network. For instance, Bahrepour et al. (2011) used the majority voting method for Parkinson patients' falling detection. The authors collected acceleration and gyroscope data from participant's feet, shank, thigh, and trunk. Sensor data, initially generated from each source, formed the preliminary detection results. These results were then transformed into a final decision via the implementation of a majority voting rule. Statistical evaluation underscored that the fusion framework offers higher accuracy than outcomes reliant on individual sensors.

### 4.4.2 Bayesian decision fusion

Bayesian inference, which is one of the widely used mathematical approaches for decision-level fusion, allows the integration of multiple pieces of evidence following the Bayesian theory of probability. In this formalism, the uncertainty of one event is represented as conditional probabilities whose values lie between zero and one. A value of zero signifies a total absence of belief, while a value of one denotes absolute certainty. The fundamental concept of Bayesian fusion involves utilizing a posterior probability hypothesis as a representation of the belief in fusion results. In specific, $P(Y|X)$ represents the posterior probability of hypothesis $Y$ given the $X$, which means the probability of event $Y$ occurring given that $X$ is true (Kendall, 1948). Assuming that $Y$ is independent of $X$, then the $P(Y|X)$ could be calculated by the Bayes rule (Equation (4-4)) with a given value of the prior probability of hypothesis $Y$, denoted $P(Y)$ (Pan et al., 1998).

$$P(Y|X) = \frac{P(X|Y)\mathrm{P(Y)}}{P(X)} \tag{4-4}$$

Likewise, the Bayesian rule allows establishing a posteriori probability under multiple conditions (e.g., $X_1$, $X_2$, …, $X_\mathrm{n}$). Expanding this framework into the multisensory problem, the decision generated from each sensor could serve as a priori probability under one condition. Therefore, the Bayesian combination of multiple data sources represents a fused likelihood that the result falls in the hypothesis. As a result, the final decision could be obtained when the fused likelihood achieves maximum value, which could be calculated with an *argmax* function (Equation (4-5)),

$$d = \mathrm{argmax}(P(Y|X_1, X_2, \dots, X_n)) \tag{4-5}$$

Despite the widespread usage of the Bayesian method, several challenges exist when utilizing the Bayesian inference in the decision-level fusion process. The first and most significant concern is the difficulty of obtaining priori probabilities, which could essentially impact the fusion reliability and performance (Steinberg & Bowman, 2017). In practice, expert knowledge is required to determine the reasonable prior probabilities with given conditions. As a trade-off scenario, setting the priori probabilities as equal, i.e., 1/n, could also be widely used when all priori probabilities are unknown. For instance, Zappi et al. (2007) put nineteen acceleration sensors in the body of workers when manufacturing automobiles. The measured acceleration signals led to training Hidden Markov Models (HMMs) for classifying ten activity classes of car production. The authors then employed majority voting and Naive Bayesian to fuse the classification results from sensors at the decision level. The classification accuracy is around 50% when using a single sensor, but adding two additional sensors increased the accuracy up to 80% by using this method. Additionally, the authors constructed a larger sensor fusion system with fifty-eight sensor nodes and achieved an accuracy of 98%. The results thereby validated that fusing additional sensors into the sensor network increases the classification accuracy of sensor networks. The test also indicated that the Naive Bayes approach performs better than majority voting in the task.

### 4.4.3 Dempster–Shafer Theory of evidence fusion

Dempster-Shafer Theory (DST) is a mathematical theory that was initially proposed by (Dempster, 1967) in the statistics domain. This theory was further developed by Shafer (1976), who introduced a mathematical object named belief function as a measurement of uncertain events rather than using possibility. Dempster-Shafer Theory is, therefore, also known as evidence theory or theory of belief. In traditional probability theories, probability is associated with a specific event. For instance, $P(A)$ represents the probability that event $A$ happened, and the probability that event $A$ does not happen is denoted as 1- $P(A)$ according to the opposite event rule. In contrast, the evidence in DST adopted the degrees of belief to measure the possibility, which could be associated with multiple events. This means that the traditional probability theory requires the probabilities for each question of interest to calculate an event's probability, while the belief functions in DST allow obtaining the belief degrees from subjective probabilities for a related one (Shafer, 1992). Therefore, the Dempster-Shafer Theory can illustrate higher-level abstraction without additional assumptions inside the evidence sets, resulting in DST as a generalization of probability theory (e.g., Bayesian theory) (Sentz & Ferson, 2002). Due to the weaker requirement of assumptions, adopting DST in fusing the results from different sensors, particularly in the multi-modality sensor system, is more flexible for generating theoretically reliable results (Castanedo, 2013). For example, Bayesian inference requires given values of priori probabilities, while the DST has no such requirement in the decision-fusion application, thereby waiving the basis caused by hand-crafted priori information (Cobb & Shenoy, 2003). In the study conducted by Chen et al. (2014), two sensor modalities were incorporated: a depth camera and an inertial sensor, aiming to enhance the Human Activity Recognition (HAR) system. The authors utilized data from the Berkeley Multimodal Human Action Database (Berkeley MHAD) by Ofli et al. (2013), which is a well-known public database containing depth images and corresponding acceleration signals of human activity. Using these data and the Collaborative Representation Classifier (CRC), an eleven-activity recognition model was trained in the research.

Moreover, two types of fusion, namely, feature-level and decision-level, were explored in Chen et al. (2014) study. The features derived from both sensor modalities were merged prior to classification when conducting the feature-level fusion. Conversely, the decision-level fusion protocol employed the Dempster-Shafer theory to integrate the outcomes from each sensor-based classifier. As per the results, an increase in activity recognition accuracy ranging between 2% and 23% was noted when compared to the performance of using individual sensors.

The Dempster-Shafer (D-S) theory has also been applied in the construction domain for multiple topics, in which risk assessment and safety monitoring are the most widely used areas.

Ding and Zhou (2013) developed a web-based risk early warning system for urban metro construction. The designed data fusion system consists of a two-stage platform. The system employs the D-S theory in the second stage to fuse independent assessments obtained from measurement, prediction, and inspection states, thereby improving the reliability and accuracy of safety risk assessment. Therefore, the proposed model can automatically make early warning decisions. Zhang et al. (2017) introduced a comprehensive method for assessing tunnel-induced risk during the initial stages of construction. This approach combines Fuzzy Matter Element (FME) analysis, Monte Carlo simulation techniques, and Dempster-Shafer (D-S) evidence theory into a hybrid framework and was validated at the Wuhan Yangtze Metro Tunnel project in China. In the case study, fourteen influential parameters were considered when using the measurements in FME to construct the Basic Probability Assignments (BPAs). In particular, the D-S theory was utilized to merge the BPAs across various risk scenarios, allowing a more reliable evaluation of safety risk awareness.

### 4.4.4 Limitations of the decision-level fusion methods

The mentioned decision-level fusion techniques in the previous sections have limitations in their applications. The majority voting rule is not functional when multiple objects are counted as the most significant or most frequent. For instance, the fusion of two sensors at the decision level is intricate if each sensor earns the same trust level. The Bayesian fusion requires the priori probabilities, which are usually absent. Either assigning the equal probability or assuming the priori value under expert knowledge is prone to import basis in calculating. In the meantime, the fundamental assumption of using the Bayesian formula is that combined entities are independent of each other. However, such an assumption is rarely valid in the applications, undermining the reliability of fusion results (Kuncheva, 2014). In addition, using the Dempster–Shafer (DS) combination to combine multiple corresponding results sometimes causes event conflict, resulting in invalid fusing outcomes (Smarandache & Dezert, 2015). In order to overcome the technique limitations of the conventional DS rule of combination, several modified rules are proposed, including Yager's rule (Yager, 1987), Dubois and Prade's (Dubois & Prade, 1988), and Murphy's rule (Murphy, 2000). The researchers also attempt to construct a comprehensive approach with multiple decision fusion methods. For instance, Sebbak and Benhammadi (2017) introduced the majority voting principle into the Dempster–Shafer Theory-based decision fusion sensor network, which was adopted in the Internet of Things (IoT)-based indoor healthcare system. The author used the simulated environment to produce the human activity signal in multiple modalities and predicted the posture into four classes (i.e., falling, sleeping, exercising, and watching TV) with different classifiers. The proposed fusion framework then processed the corresponding predictions, indicating that the merged strategy (i.e., majority voting plus Dempster–Shafer method) earned better and more intuitive classification results.

Besides the fusion technique's limitation, the difference in sensor sources should also be taken into account. In particular, the priori assumption of directly fusing sensor estimates (e.g., Dempster-Shafer combination Equation (4-13) is that each sensor produces results with the equivalent credibility. However, the assumption is constantly violated in real situations because various sensors, particularly sensors of different modalities and their associated classifiers, are hardly able to provide estimates with the same accuracy level (Ding et al., 2019). From the perspective of evidence theory, Shafer (1976) utilized the concept of discounting to explain the varying levels of trust in information sources before combining evidence. There could be situations where a support function is considered inaccurate because it does not account for uncertainties affecting the evidence. In such cases, it would be reasonable to discount the degrees of support provided by the function. When applying such a concept in the specific multisensory fusion domain, the differences in credibility between sensor estimates are the leading cause of the discounting of support degrees. Because of the inherent nature difference of sensors, they perform differently in certain circumstances. Additionally, sensors of the same type could also produce varying levels of credibility due to manufacturing differences and environmental factors such as time and temperature. Therefore, assigning the sensor estimates moderate weights is an alternative to adjusting uneven trust capability (Wu et al., 2002).

## 4.5 Methodology

This section presents a novel method for action recognition of diverse construction tasks based on both acceleration and image data by using a decision-level fusion approach, as shown in Figure 4-4. In particular, based on the previous discussion of the limitations inherent in existing decision-level fusion approaches and the biases introduced by unequal sensor trust, this study developed a comprehensive framework of weighted Demster-Shafer decision-level fusion to address these challenges. The proposed method employs the Demster-Shafer Theory as the baseline model, aiming to maximize the compatibility of the fusion framework. Meanwhile, this method also utilizes a weighting mechanism from prior knowledge on strengths of acceleration- and vision-based action recognition to balance the unequal trust problem associated with multiple sensors. This study intends to enhance the overall performance and robustness of decision-level fusion by using this comprehensive framework, thereby overcoming the previously identified limitations and biases.

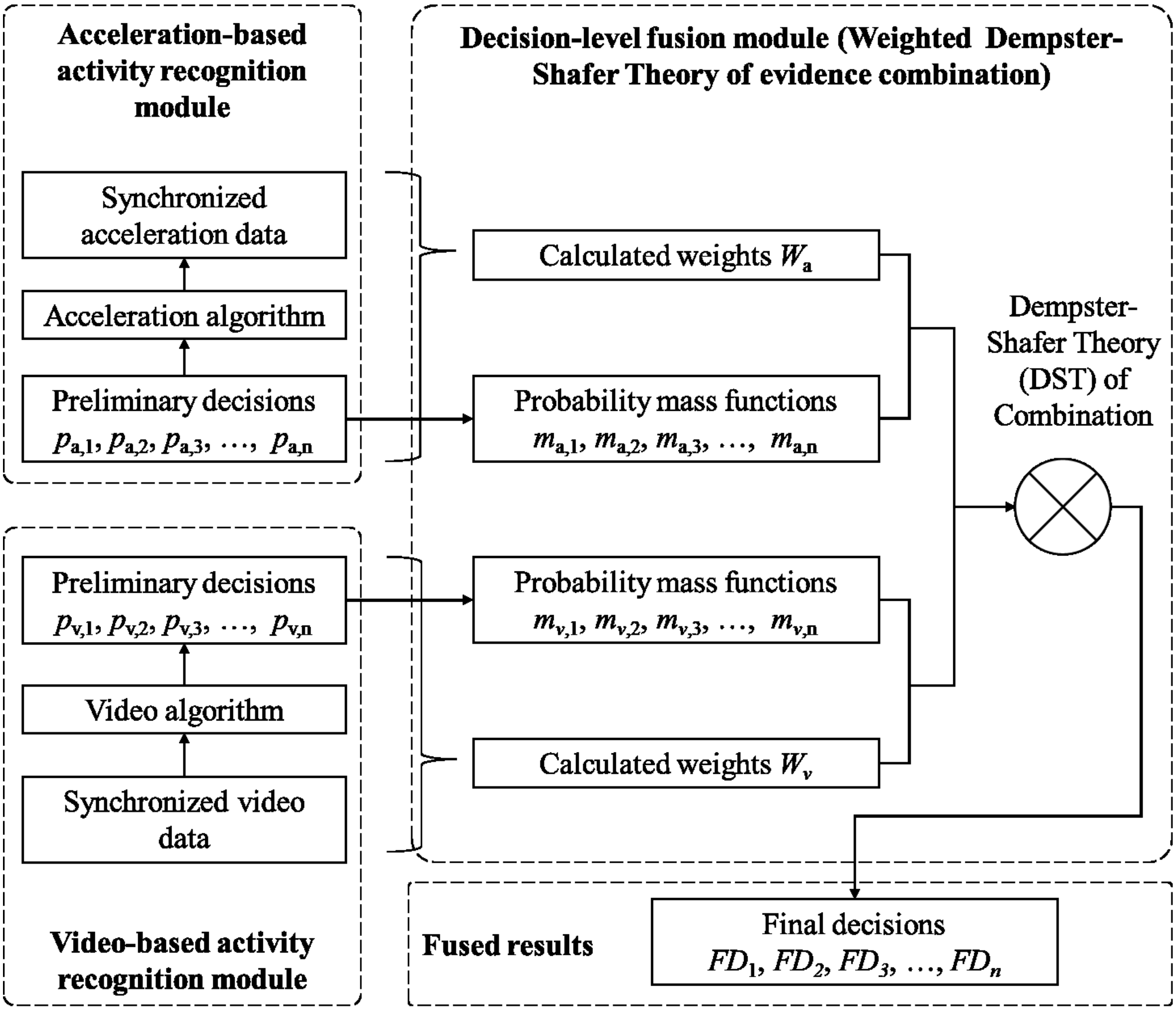

Acceleration-based activity recognition module
Synchronized acceleration data
Acceleration algorithm
Preliminary decisions
$p_{a,1}, p_{a,2}, p_{a,3}, \ldots, p_{a,n}$
Preliminary decisions
$p_{v,1}, p_{v,2}, p_{v,3}, \ldots, p_{v,n}$
Video algorithm
Synchronized video data
Video-based activity recognition module
Decision-level fusion module (Weighted Dempster-Shafer Theory of evidence combination)
Calculated weights $W_a$
Probability mass functions
$m_{a,1}, m_{a,2}, m_{a,3}, \ldots, m_{a,n}$
Dempster-Shafer Theory (DST) of Combination
Probability mass functions
$m_{v,1}, m_{v,2}, m_{v,3}, \ldots, m_{v,n}$
Calculated weights $W_v$
Fused results
Final decisions
$FD_1, FD_2, FD_3, \ldots, FD_n$

Figure 4-4 Proposed comprehensive decision-level fusion framework

### 4.5.1 Developing deep learning algorithms for action recognition

### 4.5.2 Data processing

1. Data segmentation

*Acceleration data*

When using acceleration data for Human Activity Recognition (HAR) tasks, segmentation is a necessary preprocessing step since 1) acceleration signals fluctuate over time, and segmentation could balance abnormal vibrations (Zheng et al., 2018), and 2) a complete human activity pattern shall last for a particular duration (e.g., 1.0 second). Segmenting signal sequences allows for acquiring the maximum information on human activity (Banos et al., 2014). The current study adopted the sliding window technique, a widely used signal segmentation method, to convert the consecutive time-series acceleration into smaller patterns (Dietterich, 2002). Considering the simplicity of implementation, this study segmented the acceleration data into fixed-size windows (Preece et al., 2009). A 50% overlap was also adopted when sliding the windows in order to reduce the transition noise from adjacent windows (Su et al., 2014).

The sliding window of a fixed size means that all the windows have identical durations. Since wearable sensors collect acceleration data with a constant frequency, the segmented windows are supposed to contain an equal amount of acceleration data sets. In the current study, the frequency of collecting acceleration data was set as 100 Hz in Apple Watch, and thus, one hundred data points per second were recorded. Assuming the window size is 1.5 seconds, each window would comprise 150 sets of acceleration data. However, the data hardly matched the designed numbers in one slide window due to hardware errors, such as missing storage and timestamp delay (Teh et al., 2020). Under such circumstances, the current study processed the data as follows:

1) If the counted number is less than expected, e.g., 148 data counted in a 1.5-second window, this study appended the $N$ at the tail of the window, where the $N$ refers to the number of missed data ($N$ equals 2 in this example). Specifically, the added data shall repeat the patterns seen before the added position (Datar et al., 2002);
2) If the counted number is more than expected, e.g., 152 data counted in a 1.5-second window, this study sliced the first $K$ data, where the integer number $K$ refers to the expected number of data in the selected window, and in this example, the $K$ is supposed to be 150. The extra data usually appears at the last window, as the total duration of raw data may not be able to be divided into an integer number of windows. Also, when the

time lag is significant, the data belonging to the early pattern could be shifted to the latter part, resulting in missing data in the previous window and extra data in the current window.

*Video data*

In the data preprocessing phase, the video data does not need to be segmented like the acceleration signals for learning algorithms. However, the segmentation of video data would also be needed from a data fusion point of view. Since the decision-level fusion purely combines the results from different models, the data processing of each model is not required to remain the same. However, a meaningful fused decision requires that all input decisions represent the same context, which means the data inputted to the individual model shall serve the identical event. For example, there are two decisions, *a* and *b,* which are predicted results obtained from different sensors. The requirement for applying decision-level fusion methods is that decision *a* and decision *b* represent the same event that happened at the same time. If the decisions, i.e., prediction results, are generated from time-series data sequences, the durations recorded by the two sensors should also remain identical. At the fusion phase, therefore, the video sequence is required to be segmented in order to synchronize with the acceleration data, thereby representing the identical activity event for the fusion purpose. The sliding operation towards acceleration and video data includes duration synchronization and time synchronization, which will be illustrated in the following section.

### 2. Data synchronization

Data synchronization is a processing strategy to align multiple data into a singular event, which is a necessary process before fusing the data from different sensor sources (Amundson et al., 2008). As previously discussed, video and acceleration data segmentation is compulsory before incorporating them into the decision-level fusion framework. In this regard, both timestamp synchronization and duration synchronization are applied in data processing. Firstly, an internal synchronization approach (Olson, 2010) was used to obtain a consistent time clock across two types of sensor data. Secondly, the sliding window technique was applied to divide raw data (acceleration data and videos) into uniform windows with identical sizes and overlapping settings, and the starting time of the sliding window is the unified time of both data. The synchronization process is described in Figure 4-5, the activity started at $acc\_t_{\text{start}}$ and $video\_t_{\text{start}}$ in the clock system of the accelerometer and camera clock, respectively. The window sizes of the acceleration pattern and video pattern before synchronizing are denoted as $d_{\text{acc}}$ and $d_{\text{video}}$, respectively. It is worth noting that the data for synchronization are already accurately labeled, which means the $acc\_t_{\text{start}}$ and $video\_t_{\text{start}}$ are the accurate times of activity started, which is extremely important. Otherwise, the entire synchronization system would be biased. Furthermore, in order to ensure that the labeling procedure is well-designed and operated, the detailed procedure will be illustrated in the data labeling section. Since the sensors store associated timestamps of activity in their local time systems, and the clock systems are not initially identical, $acc\_t_{\text{start}}$ and $video\_t_{\text{start}}$ may differ. The local times coordinates are transformed into a single system first, and then the timestamps from different modality sensor data are aligned. As presented in this example, the starting time of acceleration data and video are synchronized as an identical value, i.e., $t_{\text{start}}$. Then, the data streams are sliced into windows under the same size and overlapping settings. Considering the starting time of both data streams is the same, such setting of window size and overlapping could ensure that the sliced window patterns have the exact pattern sizes. This means that the synchronized acceleration and video data include the same pattern length and count, thereby ensuring the equivalent input data size and number to activity recognition algorithms.

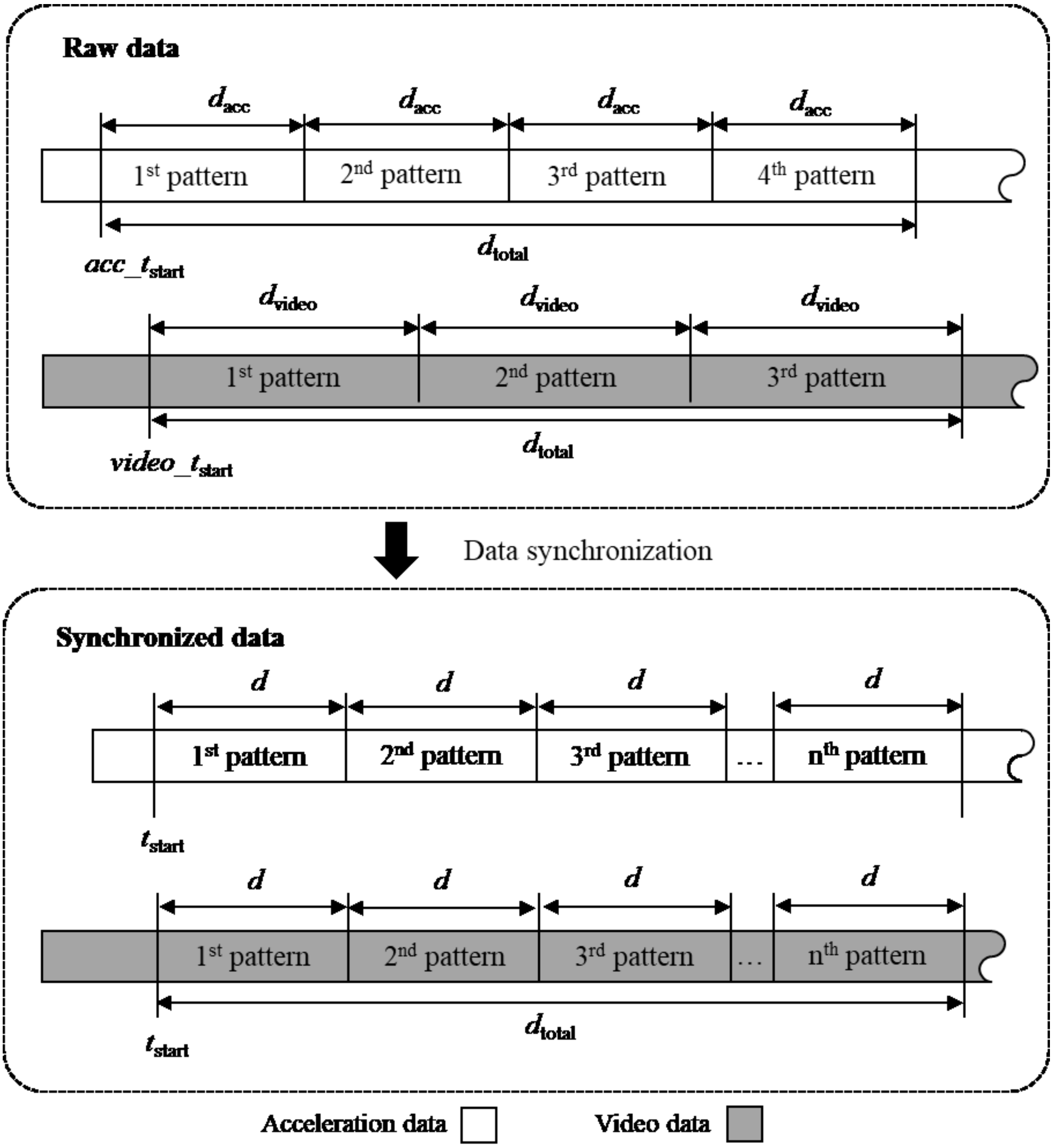


Figure 4-5 Illustration of data synchronization

*Acceleration-based action recognition algorithm*

Previous research used BiLSTM to train the action recognition classifier using acceleration data and achieved acceptable performance (Gong et al., 2022). The current study intends to utilize a similar activity taxonomy as that used in the mentioned research (i.e., Table 2-2). The identical BiLSTM algorithm described in Chapter 3 is utilized in this study.

*Vision-based action recognition algorithm*

The current study employed the ResNet framework to train the video-based action recognition classifier. The ResNet, i.e., Residual Networks, is a robust video classification architecture proposed by He et al. (2016), which was then widely used in multiple domains with a good performance, such as semantic segmentation and object detection (Dai et al., 2016; Wang et al., 2018). Though ResNet is a fundamental structure for classifying activity through videos, more advanced architectures could achieve higher detection accuracy. However, this study focused on validating the capability of the decision-level fusion method, particularly the complementarity of the fusion method. In this context, the basic architecture of the algorithm can effectively reveal the potential of the fusion method. This argument reaffirms the choice of ResNet as the foundational architecture for the current study.

ResNet is one of the most popular architectures when building a deep neural network computer vision domain because it addresses the degradation problem. In practice, stacking deeper layers is regarded as a standard option to increase network performance. However, an overly increasing network layer was found to be harmful to the network's accuracy, which is called a degradation problem (Srivastava et al., 2015). ResNet introduced residual mapping other than the underlying mapping for the layers to fit. For instance, the desired underlying mapping was denoted as $H(\mathrm{x})$, and the corresponding residual mapping is $H(\mathrm{x}) - \mathrm{x}$. Therefore, the original mapping becomes $F(\mathrm{x}) + \mathrm{x}$ when letting the stacked layers fit $F(\mathrm{x}) = H(\mathrm{x}) - \mathrm{x}$. Also, as shown in Figure 4-6 (a), the residual block in the deep residual learning framework comprises another essential component called shortcut connection, which enables bypassing the input from the block top to the tail. In particular, the identity mapping algorithm is used as a shortcut connection, which directly copies the results from the previously learned models. The purpose of shortcut connections is to ignore the additional stacked layer if the degradation issue appears, resulting in the training performance of the deeper network being no worse than the shallower one (He et al., 2016). The shortcut connections add no additional parameter to the network, thereby adding no extra computation cost. As a result, compared to optimizing layer depth, the deep residual learning framework is a cost-friendly alternative to solve the degradation problem (He & Sun, 2015).

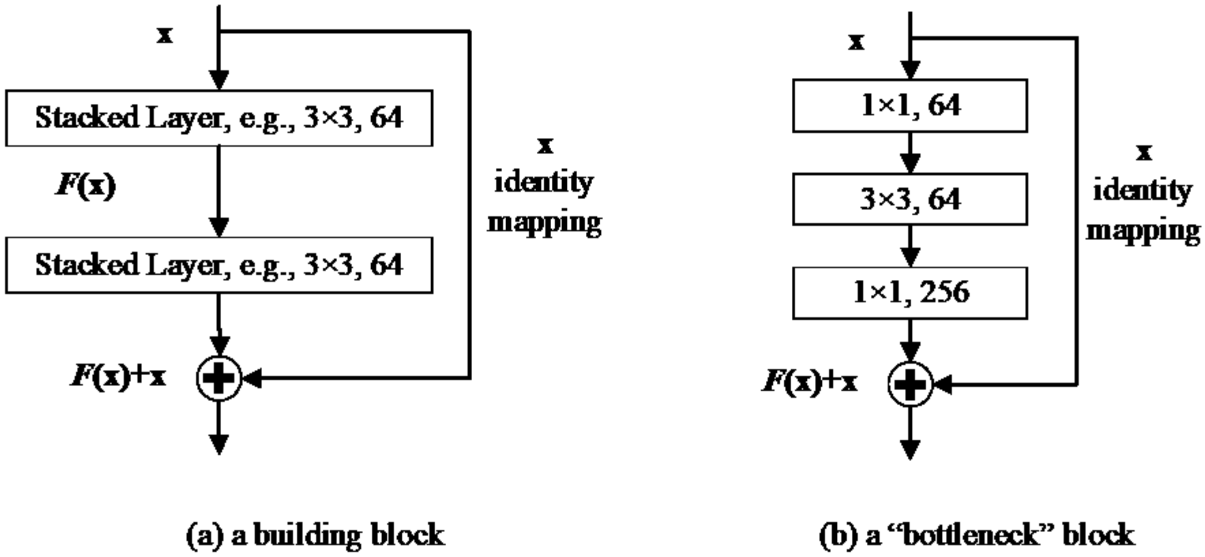


Figure 4-6 Example of residual blocks

In the ResNet architecture, the selection of superparameters, such as the depth of layers (i.e., the number of layers) and the residual block's dimensions, remain challenging. He et al. (2016) tested the ResNet structure on different recognition tasks in multiple public libraries (e.g., ImageNet, PASCAL, and MS COCO) and proposed sophisticated structures based on their test results, including 18-layer, 34-layer, 50-layer, 101-layer, and 152-layer structure, which are also known as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, respectively. The proposed ResNet-*n* structure has been widely used as the backbone of an advanced deep-learning network and has shown good generalization capability (Benali Amjoud & Amrouch, 2020). For instance, ResNet-50 has been used to develop BlitzNet (Dvornik et al., 2017) and RetinaNet (Lin et al., 2017). The ResNet-101 is employed in the structure of R-FCN (Dai et al., 2016). In this regard, the ResNet framework is regarded as suitable for the current study's video-based human activity recognition task. ResNet-50 architecture is illustrated in Figure 4-7, where a 3-layer block, named bottleneck block (Figure 4-6 b), replaces the basic block in the residual learning framework. The kernel size and kernel numbers are described as $k \times k$, $n$ in the stacked layer, such as 7 × 7, 64 in the first convolutional layer. This study also adopts a default stride size of 2 for all the stacked layers.

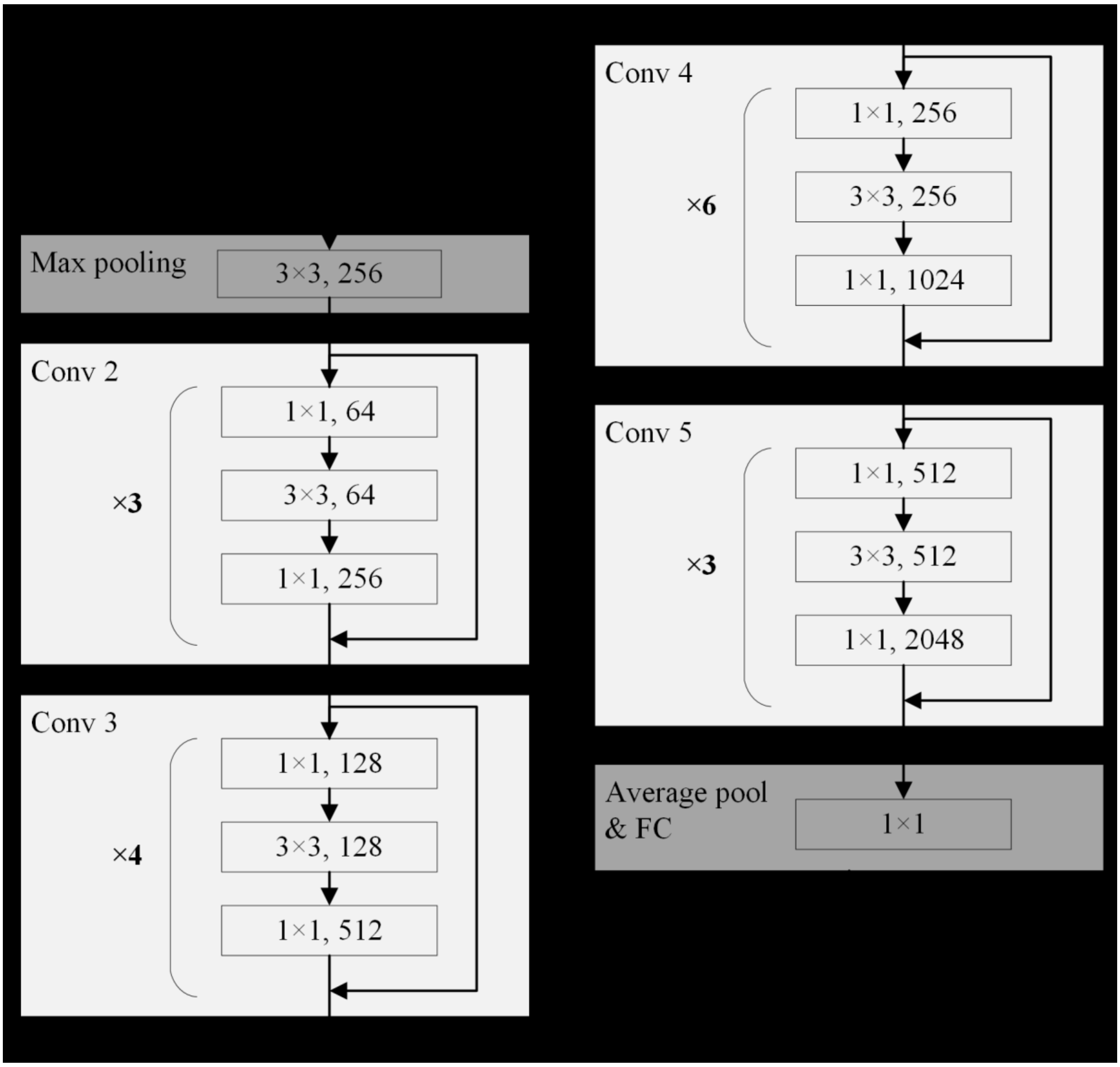


Figure 4-7 Architecture of ResNet-50

### 4.5.3 Decision-level fusion framework

In order to fuse the results predicted by acceleration-based and video-based models, the current study adopts the theory-based decision-level fusion methods that utilize the Dempster–Shafer Theory (DST) of evidence combination as the basic structure. However, to address the limitations of the existing DST fusion method, this study proposes novel weighting approaches, including 1) the Weighted Dempster–Shafer Theory (WDST) of evidence combination, 2) the Topk Weighted Dempster-Shafer (TopkWDS) method, and 3) the Thresholding Weighted Dempster-Shafer (TWDS) method.

*Fusion input*

A decision-level fusion framework has been proposed to amalgamate prediction probabilities from diverse sources, wherein the input of the fusion method originates from the acceleration-based and video-based activity recognition models corresponding to identical patterns of activity data. Within this framework, an activity recognition model trained on either acceleration or video data is expected to yield a single label as a prediction for any given input data. This structure, however, gives rise to two challenges. Firstly, the predicted label lacks probability information, which hinders the mathematical computations of the fusion method. Secondly, the scale of prediction is also a concern as the existing model produces only one label per data input, inhibiting the fusion approach from conducting a logical combination when the acceleration and video models predict differing labels for the same activity data. Hence, the final output (i.e., predicted label) from the individual sensor-based deep learning activity recognition model is unsuitable as the fusion method's input, necessitating modifications in the trained activity recognition model.

The proposed fusion method, grounded in mathematical theory, yields a final decision by generating a probability union. This process heavily relies on decimal scores drawn from participating sensor models. Therefore, targeted inputs for the activity recognition model should be those capable of producing decimal scores for each activity category within the provided dataset. Upon examining the neural network topology depicted in Figure 4-8, it becomes evident that the activated output obtained from an activation function is optimally positioned to serve as the input for the decision-level fusion method proposed in this research.

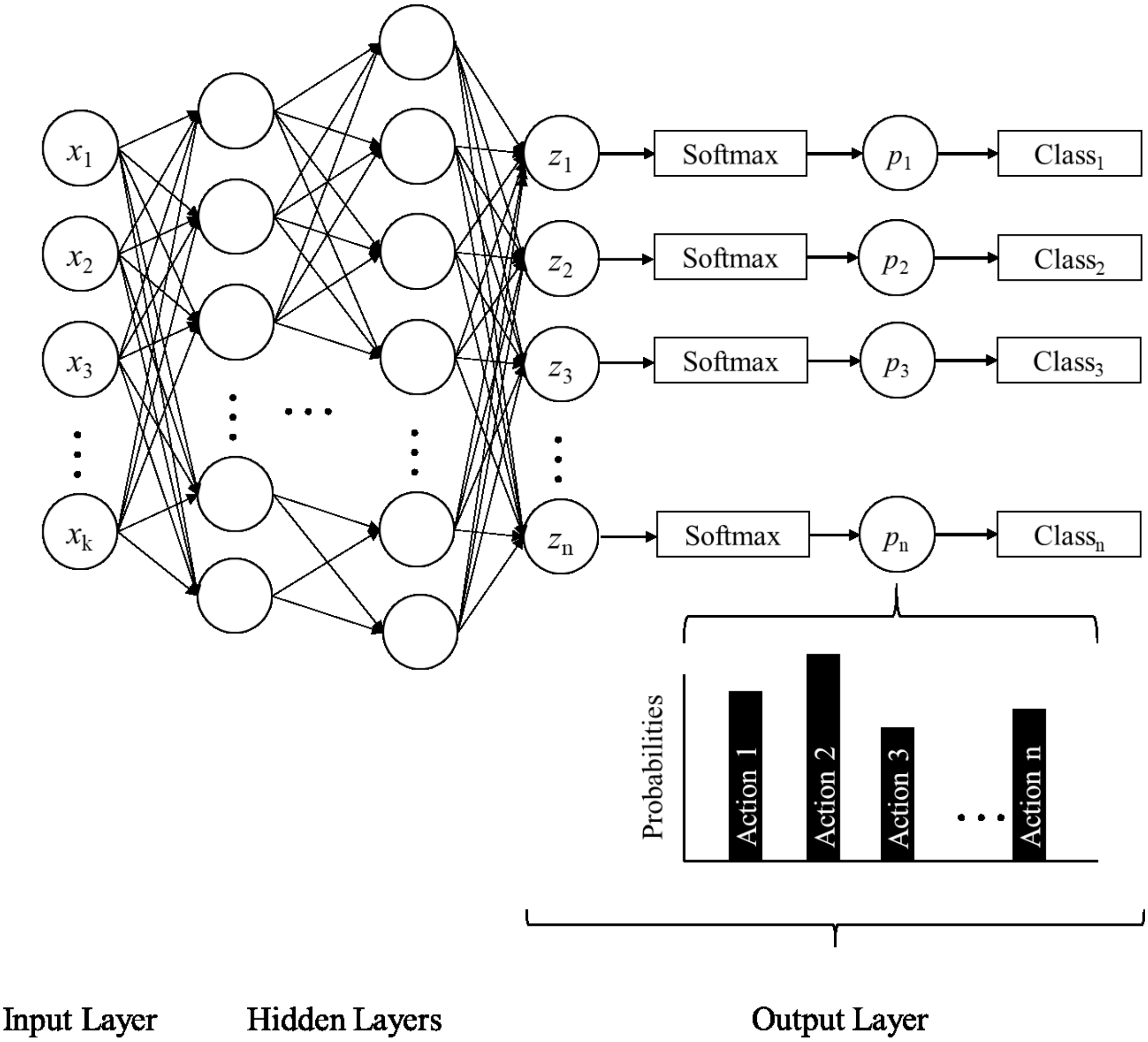


Figure 4-8 An example of neural network architecture designed for multiclass classification

The activation function in the artificial neural network serves as a transformer between layers, which translates the input signal into output information and passes it to the next layer (Sharma et al., 2017). Such transformation is particularly crucial in the output layer, where the desired predictions are derived from previous layers. Nwankpa et al. (2018) compared the popular activation functions, which include the Softmax Function, Sigmoid, Hyperbolic Tangent Function (Tah), and Rectified Linear Unit (ReLU) Function. Among those activation functions, the Softmax Function is widely used in the output layer when building a multiclass classification model (Badrinarayanan et al., 2017). As presented in Figure 4-8, this function converts the calculated amount into a probability distribution, with each number in the resulting output representing the estimated probability of a specific class given input by the model (Krizhevsky et al., 2012). It is also suggested to set the sum of decimal probabilities of prediction as 1.0 in order to speed up the convergence rate (Islam et al., 2018). The mathematical definition of the Softmax function is provided in Equation (4-6), where $\sigma$ is the Softmax function, $\vec{z}$ represents the input vector to the Softmax function, $k$ denotes the number of classes in the multiclass classification model, and $e^{z_i}$ is the exponential function of each element of the input vector $\vec{z}$. The term $\sigma(\vec{\mathrm{z}})_i$ refers to the prediction probability of $i^{\mathrm{th}}$ category in one prediction, further serving as the input for decision-level fusion methods. The fusion process is therefore could be translated as following mathematic problem: with given multiple probability sets from different sources (i.e., $P_{\mathrm{i,S1}}$, $P_{\mathrm{i,S2}}$, where $i$ represent the event number and ranges from 1 to $n$, $S_1$ and $S_2$ refers two source), combine the probabilities representing the same event into comprehensive results (i.e., $P_{\mathrm{i,S1\&2}}$). In the current study, the theory-based decision-level fusion would be denoted by the mathematical method, as illustrated in the following chapters.

The fusion process can be mathematically conceptualized as the integration of various probability sets, each derived from distinct sources. These sources are represented as $P_{i,\mathrm{S1}}$, $P_{i,\mathrm{S2}}$, where $i$ represents the event number ranges from 1 to $n$, and $S_1$ and $S_2$ correspond to two distinct sources. These multiple sets fuse to form comprehensive results depicted as $P_{i,\mathrm{S1\&2}}$. The present research employs this decision-level fusion to perform such integration, with specifics to be illustrated in the subsequent chapter.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \quad for\ i\ =\ 1, \ldots, n \tag{4-6}$$

*Fusion implementation*

The current study took the flexibility of the Dempster-Shafer Theory (DST) into account and adopted the DST's combination rule to fuse the decisions from multiple sources. Using the DST in decision-level fusion consists of two steps. The first step is to obtain the belief degrees of the target events. According to Shafer (1992), the degree of belief for one question could be estimated from the subjective probabilities of related questions. The second step is to aggregate information (i.e., degree of belief for each question) from multiple sources using Dempster's rule of combination. In the Dempster-Shafer formalism, the degree of belief is represented by the belief function or mass function. The calculation of the belief function is attached as follows (Castanedo, 2013; Sentz & Ferson, 2002):

1) Given an exhaustive and mutually exclusive frame of discernment, *T*. The power set $\Theta$ of *T* contains all possible state $\theta_i$, i.e., $\Theta = \{\theta_1, \theta_2, \cdots \theta_N\}$;
2) A hypothesis, denoted *H*, which is a subset of the power set, i.e., $\Theta$. For example, assume that $T = \{a, b\}$, then $\Theta = \{\{\emptyset\}, \{a\}, \{b\}, \{a, b\}\}$, and *H* equals to $\{\emptyset\}$, $\{a\}$ or $\{b\}$ or $\{a, b\}$. The belief degree of *H* is represented by the Basic Belief Assignment (BBA) or the mass function. The mass function meets the following requirements.

First, the empty set's mass value is zeros, as shown in Equation (4-7),

$$m(\emptyset) = 0 \tag{4-7}$$

Second, the value of the probability mass function ranges from 0 to 1, shown in Equation (4-8)

$$m: 2^{\Theta} \rightarrow [0,1] \tag{4-8}$$

Third, the masses of all the hypotheses add up to a total value of one.

$$\sum_{\{H \in \Theta\}} m(H) = 1 \tag{4-9}$$

Consider *A* as a partial component, i.e., a subset, of hypothesis *H*. Therefore, the relationship between *A* and *H* is $A \subseteq H$. The DST defines the belief of *H* as the sum of all the subset's masses, which is shown in Equation (4-10):

$$bel(H) = \sum_{A \subseteq H} m(A) \tag{4-10}$$

In the meantime, the plausibility, denoted *pl*(*H*), is defined as Equation (4-11), which represents the sum of all the masses (or BBA) of the sets *A* that intersect the set of *H*.

$$pl(H) = \sum_{A \cap H = \emptyset} m(A) \tag{4-11}$$

Moreover, the belief and plausibility represent the lower and high bounds of the probability mass function of hypothesis *H*, which is illustrated in Equation (4-12)

$$bel(H) \leq P(H) \leq pl(H) \tag{4-12}$$

By using the listed equations, the results from multiple sources could be represented as the corresponding degree of belief. Then, the next step is to aggregate the information beliefs by adopting Dempster's rule of combination, also known as joint mass (Halpern, 2017). In particular, assume $m_1$ and $m_2$ are two mass functions, and *B*, *B*', *C*, and *C*' are subsets of *H*, the combination rule is defined as Equation (4-13):

$$m_{\{1,2\}}(H) = m_1 \oplus m_2 = \frac{1}{1-K} \sum_{\{B \cap C = H\}} m_1(B) m_2(C) \tag{4-13}$$

where

$$K = \sum_{\{B \cap C = \emptyset\}} m_1(B') m_2(C') \qquad (4\text{-}14)$$

The combination also has the following properties:

$$m_{\{1,2\}}(\emptyset) = 0 \text{ and } H \neq \emptyset \qquad (4\text{-}15)$$

Following the procedures, the present study first determines the degree of belief, which is transferred from the activity recognition model's estimations. Secondly, the DST's combination rule is employed to integrate the multi-sensor estimates at a higher level, aiming to obtain results with higher confidence. Considering the practical constraints of the current application, this study formulated the following assumptions before assigning mass function value to the individual sensor's estimates:

1) The frame of discernment is constructed based on predefined activity categories. Assume the group of defined activity categories, i.e., the group of true labels in the classification model, includes "Traveling," "Lifting Brick," "Lifting Rebar," "Measuring Rebar," and "Tying Rebar". Then the frame of discernment *T* = {Traveling, Lifting Brick, Lifting Rebar, Measuring Rebar, Tying Rebar};
2) As established in the prior section, the activated result serves as the input for the fusion method (i.e., the DST combination). This outcome displays probability scores, each corresponding to a specific category within the predicted activities. However, such prediction scores lack the probabilities associated with subsets comprising multiple categories, such as {Traveling or Lifting Brick}, {Lifting Rebar, Measuring Rebar, or Tying Rebar}. Thus, the prediction model is limited to single categories only, and the probability estimation, such as either Lifting Rebar or Measuring Rebar, is deemed invalid under this assumption. Therefore, the hypothesis space *Θ* is restricted to singleton subsets exclusively derived from *T*.

In the meantime, the null or void set (i.e., $\varnothing$ or $\{\}$) is excluded in this scenario because the designed algorithm for multi-class classification is guaranteed to produce a single predicted class as the output. Based on the assumptions, the hypothesis space in this example has been formulated, consisting of five situations enumerated as follows:

$$\Theta = \{\{\text{Traveling}\}, \{\text{Lifting Brick}\}, \{\text{Lifting Rebar}\}, \\ \{\text{Measuring Rebar}\}, \{\text{Tying Rebar}\}\} \tag{4-16}$$

After constructing the hypothesis space, the subsequent task involves assigning belief degrees derived from the prediction probabilities of activity classes obtained from the neural network through the individual sensor estimations. Initially, the mass function of the estimations from the accelerometer and camera are defined as $m_\mathrm{a}$ and $m_\mathrm{v}$, respectively. Next, as stated in the data processing section, the complete dataset, encompassing both acceleration and videos, is divided into uniformly sized windows for classification. As a result, each data window corresponds to a distinct set of prediction probabilities and, consequently, a corresponding mass function. Therefore, for a given data window, a set of probabilities $p_i$ is generated and subsequently designated as the mass function $m_\mathrm{i}$, where $i$ ranges from 1 to $n$, indicating the index of data windows. Thus, the mass functions from the accelerometer and video sources for the $i^{\mathrm{th}}$ data window are represented as $m_{a,\,i,}$ and $m_{v,\,i}$, respectively.

Consider a specific pattern of acceleration data, $acc_\mathrm{i}$, as an instance. It is input into the acceleration-based action recognition algorithm (Figure 3-3), which produces a preliminary decision using the activation function in the output layer shown in Figure 4-8. Considering a five-category taxonomy, "Traveling," "Lifting Brick," "Lifting Rebar," "Measuring Rebar" and "Tying Rebar," Considering a five-category taxonomy Equation (4-6). This formulation yields the prediction probability series: $P_\mathrm{i} = \{p_{\mathrm{Traveling,a,i}}, p_{\mathrm{Lifting\ Brick,a,i}}, p_{\mathrm{Lifting\ Rebar,a,i}}, p_{\mathrm{Measuring\ Rebar,a,i}}, p_{\mathrm{Tying\ Rebar,a,i}}\}$. Following this, the probabilities set is employed for determining the corresponding data pattern's mass function value, denoted as $m_{\text{a,i}}$,. Specifically, $m_{\text{a,i}}$, is represented by $\{m_{\text{Traveling,a,i}}, m_{\text{Lifting Brick,a,i}}, m_{\text{Lifting Rebar,a,i}}, m_{\text{Measuring Rebar,a,i}}, m_{\text{Tying Rebar,a,i}}\}$. By performing the same framework, the mass function of each video data and acceleration data pattern can be calculated and denoted as $m_{\text{a,i}}$, and $m_{\text{v,i}}$, respectively, where *i* is the index of the data pattern and ranges from 1 to *n*.

The next step is to integrate the estimates from various sources. Specifically, the mass functions from two sources can be combined using the Demspter-Shafer rule of combination (Equation (4-13) and (4-14)) when they correspond to the same event. For this purpose, strict synchronization of the acceleration and video data is necessary to ensure that the combined components $m_{\text{a,i}}$, and $m_{\text{v,i}}$ occur in the identical timestamps of the unified timeline. Then, with a given pair of $m_{\text{a,i}}$ and $m_{\text{v,i}}$, the combined degree of belief is calculated as follows:

$$m_{\{a,v\},i}(A) = m_{a,i} \oplus m_{v,i} = \frac{1}{1-K}\sum_{E_a \cap E_v = A} m_{a,i}(E_a) m_{v,i}(E_v) \tag{4-17}$$

$$K = \sum_{E_a{}' \cap E_v{}' = \emptyset} m_{a,i}(E_a{}') m_{v,i}(E_v{}') \tag{4-18}$$

where $E_{\text{a}}$ and $E_{\text{a}}$' represent the evidence observed by the accelerometer. $m_{\text{a, i}}(E_{\text{a}})$ and $m_{\text{a, i}}(E_{\text{v}}')$ refer to the associated mass function for $E_{\text{a}}$ and $E_{\text{a}}$'. Similarly, the $E_{\text{v}}$ and $E_{\text{v}}$' mean the camera's observations, and their corresponding mass functions are denoted as $m_{\text{v,i}}(E_{\text{v}})$ and $m_{\text{v,i}}(E_{\text{v}}')$. In addition, *A* stands for the intersection of the hypothesis that combined the accelerometer and the camera's observations. Note that *A*, $E_{\text{a}}$, $E_{\text{a}}$', $E_{\text{v}}$, and $E_{\text{v}}$' are all the subsets of the hypothesis space *Θ*, representing our case's predicted construction activity category. For instance, in order to determine the belief degree of the predicted activity being "Tying Rebar" by combining the acceleration and video estimations of $i^{\text{th}}$ data pattern, the mass function of $m_{i,\{a,v\}}$(Tying Rebar), needs to be calculated.

Per Equations (4-17) and (4-18), the calculation of joint mass function requires the summation of all possible combinations that satisfy. $E_a \cap E_v =$ Tying Rebar, and $E_a{}' \cap E_v{}' = \emptyset$. When the hypothesis space $\Theta$ is the power set of $T$, each sensor observation can map to multiple predictions, for example, {Tying Rebar, Measuring Rebar}, meaning the predicted activity is Tying Rebar or Measuring. Therefore, $E_a \cap E_v =$ Tying Rebar could consist of various situations, such as $E_a$ = {Tying Rebar, Lifting Brick} and $E_v$ = {Tying Rebar}. Similarly, $E_a \cap E_v = \emptyset$ also has multiple alternatives, such as $E_a$ = {Lifting Brick} and $E_v$ = {Traveling, Idling}. The current study has assumed that each hypothesis is a singleton subset and exclusive to other hypotheses. In this regard, $E_a \cap E_v =$ Tying Rebar only contains unique situations: $E_a$ = {Tying Rebar} and $E_v$ = {Tying Rebar}, i.e., $E_a = E_v = A$. In the meantime, $E_a$' $\cap E_v$' $= \emptyset$ indicates that the accelerometer and camera produce different readings. As a result, Equation (4-17) can be simplified in this case. Consider

$$m_{i,\{a,v\}}(A) = m_{a,i} \oplus m_{v,i} = \frac{1}{1-K} m_{a,i}(A) m_{v,i}(A) \tag{4-19}$$

$$K = \sum_{E_a{}' \cap E_v{}' = \emptyset} m_{a,i}(E_a{}') m_{v,i}(E_v{}') \tag{4-20}$$

The singleton assumption employed in the current study more accurately reflects the reality of the monitored construction activity, thereby streamlining calculations and enhancing the study's applicability to real-world contexts.

The framework of the Dempster–Shafer fusion method is shown in Figure 4-9. The $p_{v,1}$, $p_{v,2}$, …, $p_{v,n}$ refer to the probability set of video data 1, 2, …, $n$ gained from the video algorithm. So, the $p_{v,n}$ is a one-dimension vector whose length equals the total number of classes. Likewise, the probability sets from the acceleration algorithm are denoted as $p_{a,1}$, $p_{a,2}$, …, $p_{a,n}$, where the numbers 1, 2, …, $n$ in the subscript represent the data number. Since the video and acceleration data are synchronized before being inputted into its action recognition algorithm, the obtained probability set of $i^{\text{th}}$ acceleration data, i.e., $p_{\text{a,i}}$ serves the same hypothesis as $p_{\text{v,i}}$ do, thereby the $p_{\text{a,i}}$ and $p_{\text{v,i}}$ could be fused to $Bel_{\text{i}}$ without additional adjustment. After balancing the fused values, the final decision would be determined and denoted as $FR_{\text{i}}$. The presented DS approach served as the baseline method in the current study.

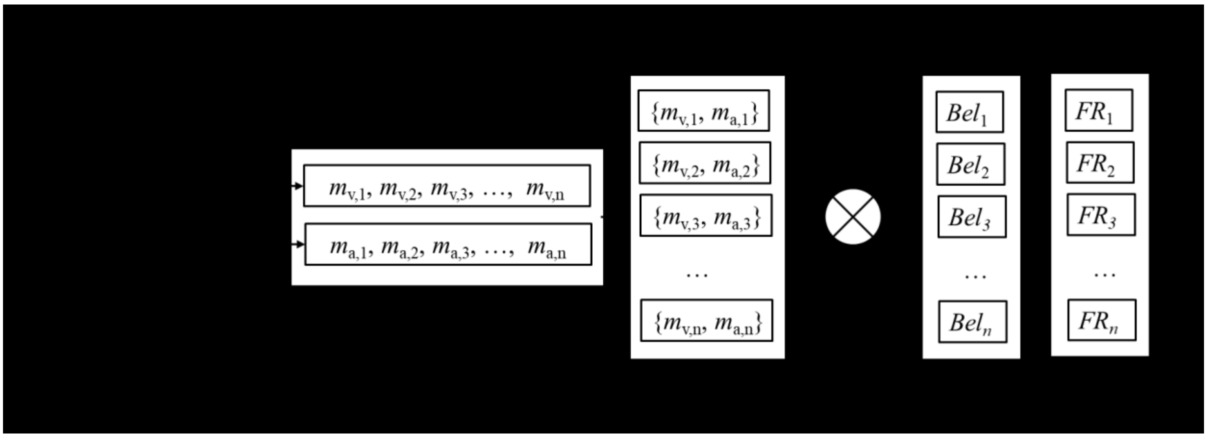


Figure 4-9 Dempster–Shafer decision-level fusion framework

*Method 1 - Weighted Dempster–Shafer (WDST) method*

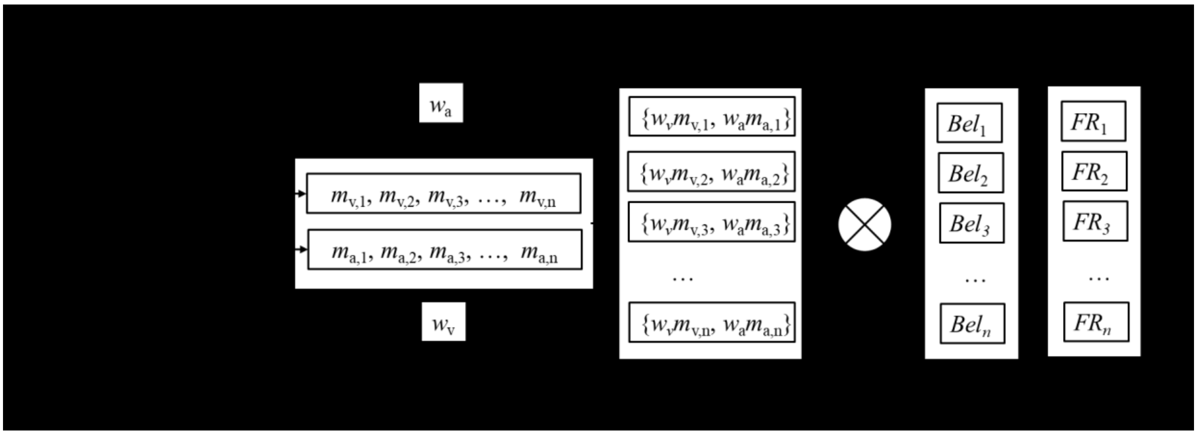
The fusion techniques themselves introduce certain biases to the fusion performance. Another notable source of fusion error is caused by the disparate credibility levels associated with different information sources, i.e., sensors and their corresponding algorithms. The current research adopts a discounting approach that assigns weights to balance various sources' influence, thereby effectively addressing the unequal credibility issue in the sensor system.

Since the Dempster-Shafer Theory-based combination approach is applied in the study as a baseline model to combine the results from different sources, the modified method is named the Weighted Dempster–Shafer Theory (WDST) method after introducing the discounting variables. This approach suggests evaluating the sensor performance in the historical data and calculating the weights, which are then assigned to the sensor result for balancing the trust level.

Hence, the previous combination formula, i.e., Equation (4-19), is modified to the following formula:

$$m_{i,\{a,v\}}(A) \ = m_{a,i} \oplus m_{v,i} = \ \frac{1}{1-K} w_{a,i} m_{a,i}(A) \times w_{w,i} m_{v,i}(A) \qquad (4\text{-}21)$$

Where $m_1(B)$ is the observation from sensor 1, $m_2(C)$ is the observation from sensor 2, and $w_1$ and $w_2$ represent the calculated weights for sensors 1 and 2, respectively. The weight of each sensor corresponds to a series of discounting amounts for each category, representing the credibility of the sensor model in classifying that particular category of activity. With a given prediction, the decimal probability of each category is assigned its corresponding discounting variable. These weighted values are then fused using the DS combination rule, the procedure of which is visually represented in Figure 4-10.



$w_a$
$m_{v,1}, m_{v,2}, m_{v,3}, \ldots, m_{v,n}$
$m_{a,1}, m_{a,2}, m_{a,3}, \ldots, m_{a,n}$
$w_v$
$\{w_v m_{v,1}, w_a m_{a,1}\}$
$\{w_v m_{v,2}, w_a m_{a,2}\}$
$\{w_v m_{v,3}, w_a m_{a,3}\}$
…
$\{w_v m_{v,n}, w_a m_{a,n}\}$
$\otimes$
$Bel_1$
$Bel_2$
$Bel_3$
…
$Bel_n$
$FR_1$
$FR_2$
$FR_3$
…
$FR_n$

Figure 4-10 Weighted Dempster–Shafer decision-level fusion framework

As presented in the weighted method formula, the performance of the proposed weighted approach relies on the weights' reliability. In order to obtain reasonable weights, Wu et al. (2002) suggested extracting such correctness variables from the historical data, and the more similar the sensor application situation to the current circumstance, the more accurate the weight. This study utilizes information obtained from the training process and prior knowledge to derive appropriate weights for fusing sensor estimates rather than replicating similar experiments to collect historical data. The current study splits the whole dataset into the training, validation, and testing data for classifier training and testing purposes. The model was trained using the training data during the training process, and the validation dataset was employed to assess training performance and prevent overfitting. In this regard, validation accuracy indicates a model's actual performance more accurately than training accuracy. Consequently, the current study selects the validation performance as the weight for sensor estimates.

When selecting weights for machine learning classification models, it is essential to consider accuracy and classification metrics such as precision, recall, and F1 score. The definitions and calculations of these metrics are delineated from Equation (3-1) to Equation (3-3). Each metric provides insight into classifier performance: Precision calculates the accuracy of positive predictions against total positives; recall evaluates the classifier's ability to identify all actual positives; and the F1 score balances precision and recall, which are crucial in imbalanced class distributions. The current study aims to examine the influence of varying weight selections on fusion outcomes through the analysis of all three metrics.

*Method 2 – Topk Weighted Dempster-Shafer (TopkWDS) method*

The examination indicates that prediction scores for categories are not evenly dispersed. A significant portion of prediction scores for a category tends towards extremely low values. For instance, classification scores corresponding to categories "Drilling," "Tying rebar," "Lifting rebar," "Measuring rebar," "Hammering," "Idling," "Lifting brick," and "Traveling" are {0.998, $1.267\times10^{-3}$, 2.477×10-4, $1.723\times10^{-4}$, $9.686\times10^{-5}$, $4.458\times10^{-5}$, $2.257\times10^{-5}$, $4.172\times10^{-6}$}, respectively These extremely small values represent a negligible chance of classification for the corresponding categories. Such small values would not significantly affect the classification results when using a single sensor source. However, in the DS combination process, the decimal possibilities of each category significantly would have an important role in the weighted method calculation according to Equation (4-21). For instance, both $10^{-5}$ and $10^{-7}$ indicate a low probability of being predicted as a particular category. They have a substantial difference of two orders of magnitude when the weights are incorporated into the formula. To mitigate such an impact, one approach is setting the value to zero. However, the implementation of activation functions, such as the Softmax function, prevents the prediction score from reaching zero. To address the minimal value impact in the weighting process, the current study proposes filtering the weights before utilizing them to adjust the prediction reliability in the decision-level fusion procedure. Inspired by the concept of top-k accuracy, the study introduces a filtering weight method called top-k weighted selection. This method involves sorting the possibilities in descending order and selecting the k largest possibilities as the primary discounting components. In the given example, if k is set to 5, the filtered prediction scores would be

$$\{0.998, 1.267\times10^{-3}, 2.477\times10\text{-}4, 1.723\times10^{-4}, 9.686\times10^{-5}\}$$

representing the decimal possibilities for the predicted categories "Drilling," "Tying rebar," "Lifting rebar," "Measuring rebar," and "Hammering," respectively.

Additionally, the DS theory requires that all mass functions sum to 1. The study suggests combining the dumped possibilities into one non-singleton set to represent the dumped category to fulfill this requirement. The prediction score would be the sum of the dumped values. In this case, the filtered prediction example after using the top-k method would be

$$\{0.998, 1.267\times10^{-3}, 2.477\times10^{-4}, 1.723\times10^{-4}, 9.686\times10^{-5}, 7.082\times10^{-5}\}$$

The corresponding prediction category set would be

{{Drilling}, {Tying rebar}, {Lifting rebar}, {Measuring rebar}, {Hammering}, {Idling, Lifting brick, Traveling}}.

The filtered weighted would then be inputted into the discounting process, as shown in Figure 4-10.

*Method 3 – Thresholding Weighted Dempster-Shafer (TWDS) method*

The Thresholding Weighted Dempster-Shafer (TWDS) method is proposed to mitigate the impact of extreme weight values. Similar to the Topk method discussed earlier, the TWDS method also arranges the prediction scores in descending order and filters out weights below a given threshold, such as 0.99. The remaining possibilities and corresponding prediction categories are combined into a subset, following the TopK method. For instance, the prediction scores for the categories "Drilling," "Tying rebar," "Lifting rebar," "Measuring rebar," "Hammering," "Idling," "Lifting brick," and "Traveling" are {0.998, 1.267×10$^{-3}$, 2.477×10-4, 1.723×10$^{-4}$, 9.686×10$^{-5}$, 4.458×10$^{-5}$, 2.257×10$^{-5}$, 4.172×10$^{-6}$}, respectively. With a given thresholding of 0.99, the filtered result should be:

$$\{0.998, 1.267\times10^{-3}, 5.889\times10^{-4}\}$$

The corresponding prediction category set would be

{{Drilling}, {Tying rebar}, {Lifting rebar, Measuring rebar, Hammering, Idling, Lifting brick, Traveling}}.

The weights obtained directly from the validation process will be replaced with the filtered values, which will then be inputted into the discounting process, as shown in Figure 4-10.

## 4.6 Data Collection Through Experimental Settings

### 4.6.1 Testing conditions

*Experimental settings*

This study, aiming to simulate the construction activity conducted on sites, deployed entities (materials, equipment, and workers) akin to those found on construction sites. In particular, a 2.4m (length) × 0.9m (width) × 0.9m (height) rebar form was placed in the testing lab. The participants were asked to conduct designed activities in the surroundings of the specimen. In this regard, the background of collected data, particularly the videos, is noisy. Though the noise makes building well-performed models difficult, the data collected in such experiment settings are closer to the raw data from the construction sites. The photos of the testbed can be found in Figure 4-11, and the layout is attached in Figure 4-12.

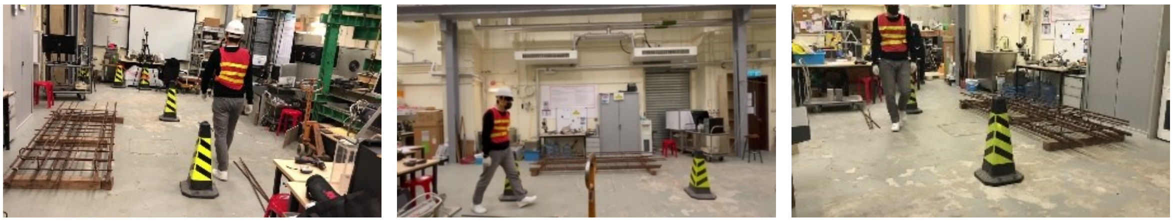


Figure 4-11 Photos of the testbed (right side view, front view, and left side view)

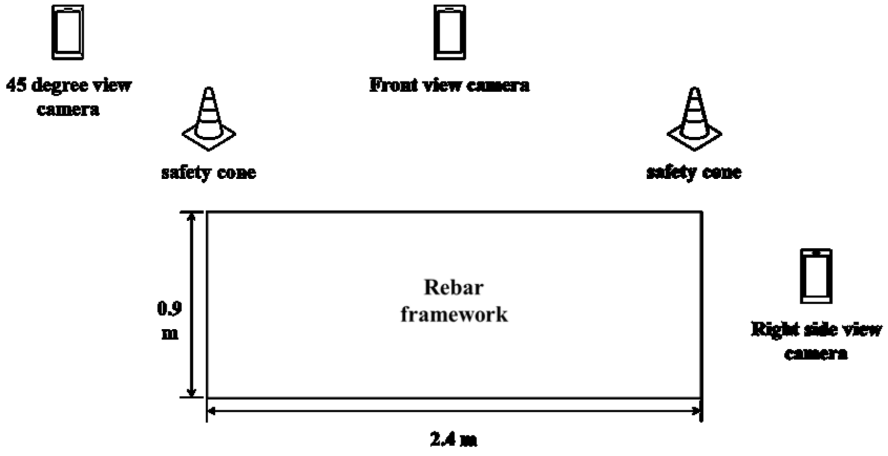


Figure 4-12 Experiment layout

*Participants*

In the experiment, three research personnel (two male and one female) from The Department of Building and Real Estate, Hong Kong Polytechnic University, were recruited to perform the designed construction activities. When experimenting, the participants are asked to wear an entire set of safety gear (i.e., safety vest, helmet, and boots) during the whole testing period (Figure 4-13) in order to 1) keep the participant safe and 2) simulate the real construction environment.

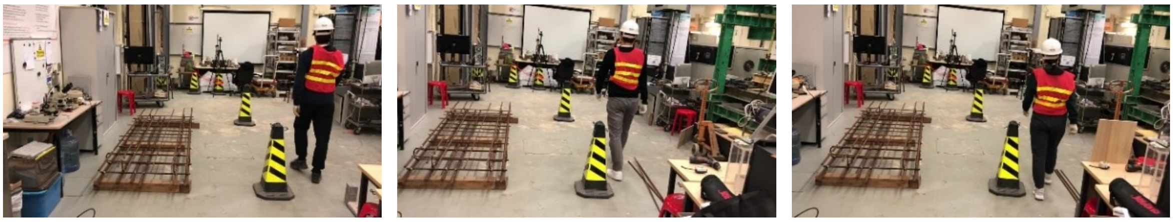


Figure 4-13 Right-side view photos of participants 1, 2, and 3 in the test

### 4.6.2 Data collection procedures

<u>*Acceleration measurements*</u>

The current study adopted the Apple Watch as the acceleration data collection sensor, which is widely applied in obtaining human activity data (Ashry et al., 2020; Kwon & Choi, 2018). In this experiment, one Apple Watch (Series 4, 40mm Aluminum & Ceramic Case) was attached to the participant's dominant hand (Figure 4-14) to collect the acceleration signals. The size of the watch is 40mm (height) × 34mm (width) × 10.7mm (depth), and the weight is 30.1g, as instructed. The researcher group also developed a data collection App in the watchOS platform to collect, store, and transfer the acceleration signal (Figure 4-14). When the experiment starts, the user is supposed to activate the embedded Inertial Measurement Unit (IMU) through the App and record the three-axis acceleration signals. The data are stored in the Apple Watch's hard drive with its operation timestamp. Meanwhile, a wireless module in the developed App will then help transfer the stored data to a paired machine, such as an iPhone or desktop. In order to minimize the data bias caused by the device, participants used the same Apple Watch in the entire experiment. The selected watch was initialized and tested as functional before the tests, and no further calibration is needed.

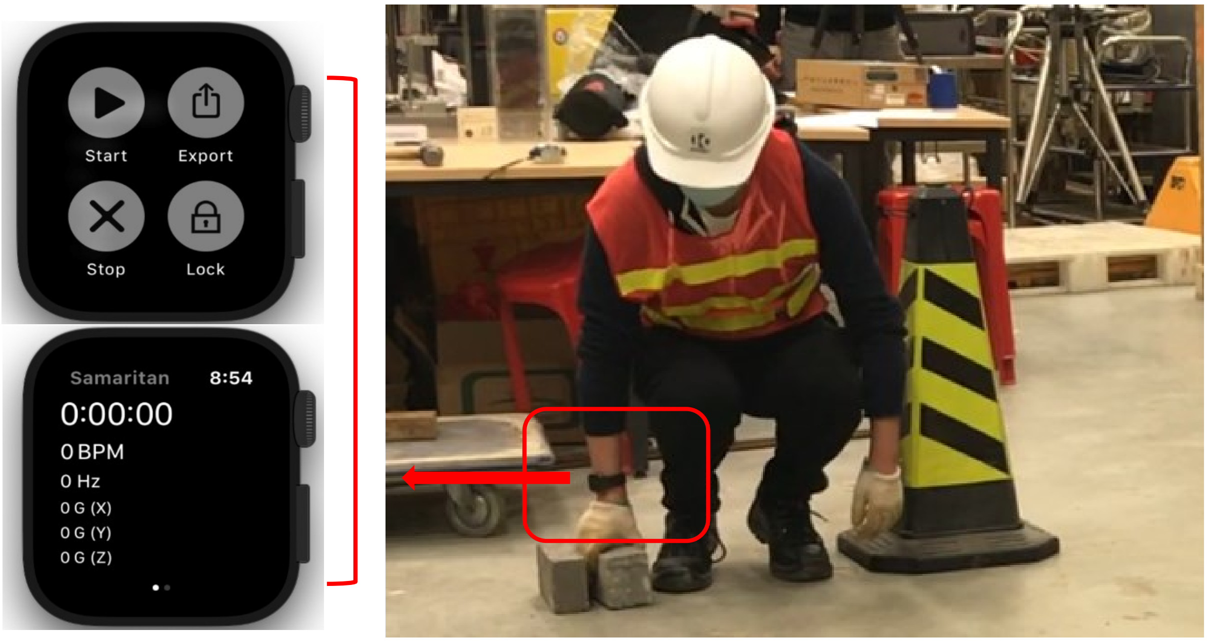


Figure 4-14 The Apple Watch placement (**right**) and data collection interface (**left**)

*Video measurements*

In the meantime, the current study used three iPhones to record the videos because 1) it is portable and flexible to install in designed positions; 2) it is a widely used commercial-grade device that also supports recording high-resolution video; 3) it brings convenience and consistency of data processing when collecting both video data and acceleration data in the same platform (i.e., Apple platform) (Chen et al., 2021). As shown in Figure 4-12, the smartphones are deployed in three positions so that the researchers can simultaneously collect the videos (i.e., front view, side view, and 45-degree view) from three points of view. Meanwhile, each smartphone is installed on tripods, and tripod height or position modification is not allowed during the whole experiment. In this regard, the videos collected in the investigation are strictly from fixed spots.

*Experiment design and procedure*

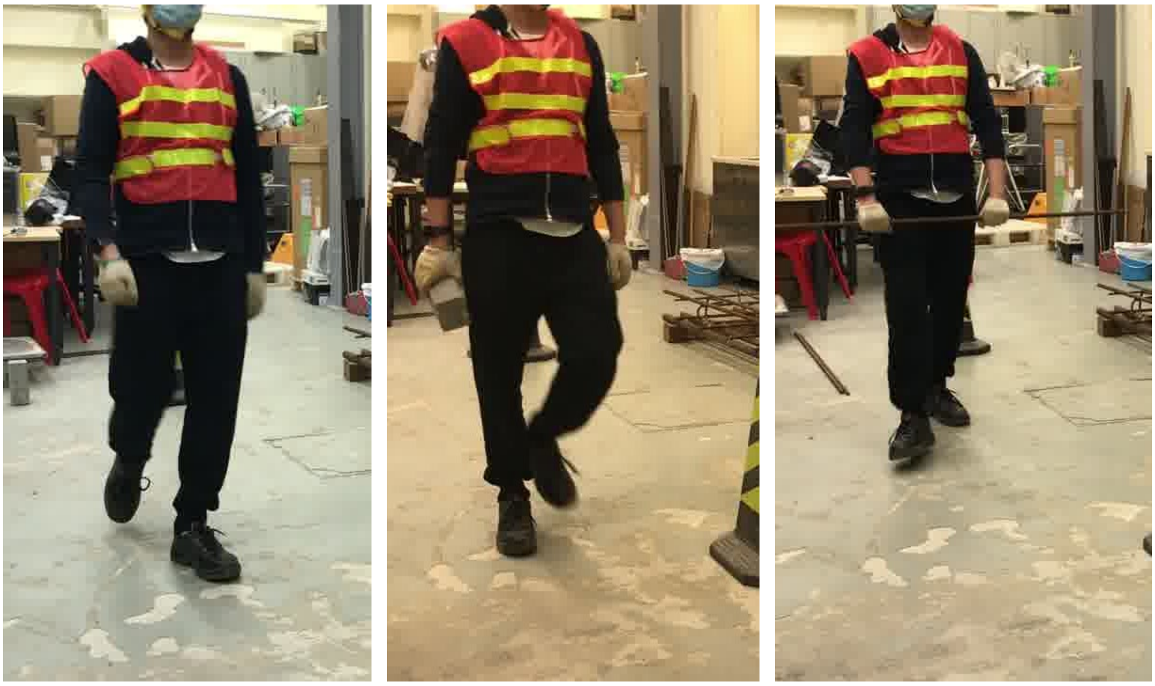
In order to validate the sensor fusion approach in construction activity for developing a more effective and stable framework of worker monitoring, typical activities from essential construction tasks are designed and conducted in the laboratory. Inspired by the proposed taxonomy in 0, the experiment designed eight different types of construction activities from the group of Level 3 Activity. Detailed information can be found in Table 4-2.

Table 4-2 Activity taxonomy used in the decision-level fusion experiment

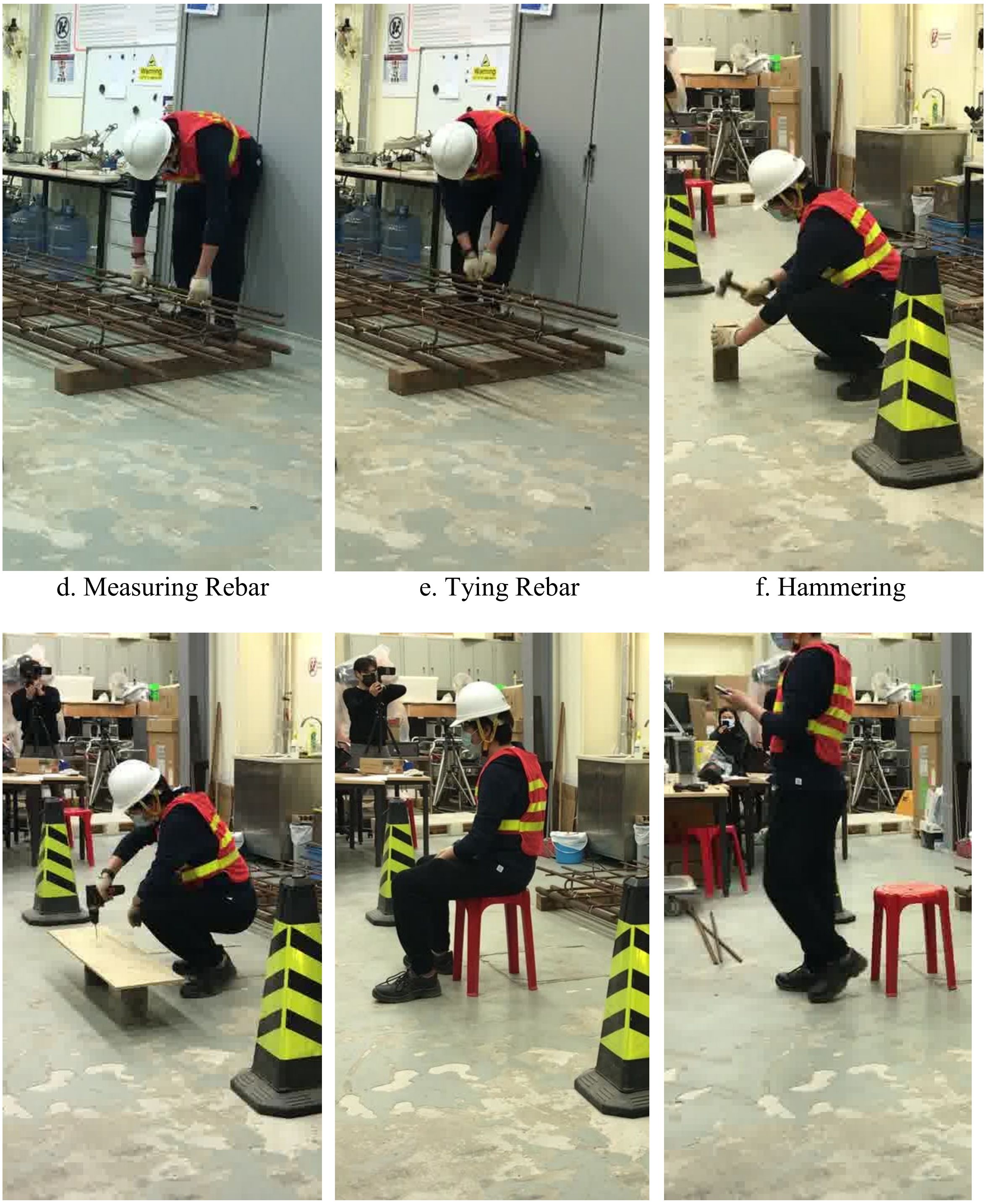
| **Activity ID** | **Level 3 Activity** | **Level 2 Activity** |
|---|---|---|
| 1 | Traveling | Traveling |
| 2 | Lifting Brick | Effective work |
| 3 | Lifting Rebar | Effective work |
| 4 | Measuring Rebar | Effective work |
| 5 | Tying Rebar | Effective work |
| 6 | Hammering | Effective work |
| 7 | Drilling | Effective work |
| 8 | Idling | Non-effective work |

The current study, therefore, designed eight different types of construction activities (Figure 4-15): “Traveling,” “Lifting Brick,” “Lifting Rebar,” “Measuring Rebar,” “Tying Rebar,” “Hammering,” “Drilling,” “Idling.” The “Traveling” required the participants to walk around the rebar specimen at a consistent speed (Figure 4-15 a). Also, the participants were required to carry brick and rebar when conducting such circus walking in order to perform the activities “Lifting Brick” (Figure 4-15 b) and “Lifting Rebar” (Figure 4-15 c), respectively. In addition, “Measuring Rebar” and “Tying Rebar” required the participants to measure the rebar grid (Figure 4-15 d) and tie the steel wire (Figure 4-15 e) along the rebar specimen, which contains abundant hand movement, resulting a longer duration compared to the pure movement jobs (i.e., “travel”) and moving materials jobs (i.e., “Lifting brick” and “Lifting Rebar”). Apart from the activities conducted around the rectangular rebar form, “Hammering,” “Drilling,” and

“Idling (including sitting and standing)” were performed in a constrained area in front of it. “Hammering” involved repeat knocking activity on the brick (Figure 4-15 f). The “Drilling” required participants to drill holes in the wooden slab by using the drilling machine (Figure 4-15 g). The “Idling” required to keep stationary in the spot. As shown in Figure 4-15 h and Figure 4-15 e, the “Idling” consists of two types of action, i.e., “Sitting” and “Standing”. Participants were allowed to make minor physical adjustments like swinging a leg or raising a hand but not aggressive movements like jumping or standing up to maintain consistency and accuracy in the experiment.



a. Traveling b. Lifting Brick c. Lifting Rebar



d. Measuring Rebar

e. Tying Rebar

f. Hammering

g. Drilling

h. Idling (Sitting)

i. Idling (Standing)

Figure 4-15 Photos of performing designed construction activity

The current study adopted a repeated measures design in a single test. As shown in Figure 4-16, each participant was required to repeat each activity five times in a continuous sequence. After finishing all five actions, the participants returned to the starting spot, which is a fixed position for the entire experiment. The researchers helped deactivate the data collection app in Apple

Watch, stopped video recording from three iPhones, and then stored the collected data (i.e., both acceleration data and video data) in the hardware. In order to minimize data collection bias, the examination of sensor positions is necessary before starting the data collection of the next activity. Specifically, the Apple Watch is required to be attached to the exact position of the wrist.

The direction of the watch is also placed in the same direction so that the collected acceleration data share the same reference coordinate system. Also, the position and height of tripods remain constant to ensure the videos are recorded from a fixed point of view. Then, the researchers re-activated the data collection function in smartwatches and smartphones to collect the motion data of the subsequent construction activity.

Before starting the experiment, the participants were required to observe the representative activity performed by the researcher for a better understanding of the designed activities. Thereby, participants could perform the same activity similarly. The experimental procedures (Figure 4-16) were instructed to the participants before testing. When executing the experiment, the participants were asked to perform tasks at their own comfortable pace. They were also suggested to take a break after finishing the data collection of one category of activities in order to reduce fatigue. During the break, the researchers stored the data and prepared the device for the next activity data collection session.

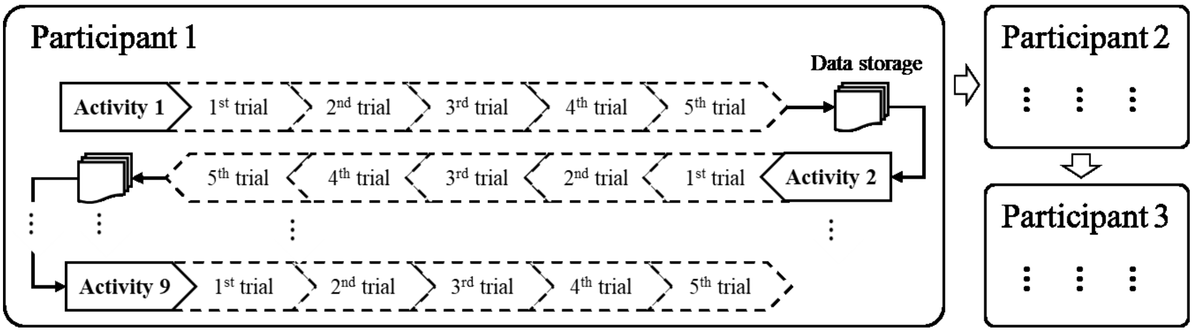


Figure 4-16 Repeated measure design of simulated experiment

*Data labeling*

As instructed in the experimental procedure (Figure 4-16), the five-trial data of one activity are stored before performing the next one, resulting in instant data labeling for both acceleration and video data. For instance, after Participant No.1 finished the fif$^{th}$ trial of Activity No.1, the researchers were supposed to help stop the video and acceleration recording. In the meantime, the acceleration data were stored in the Apple Watch's local memory, and the filmed videos were stored in the photo album of the iPhones. In Apple Watch's platform, the event name (i.e., activity 1) associated with the participant ID (i.e., Participant No.1) is included in the name of the stored file. Likewise, the event and operator information of videos could also be typed into the video instructions in the iPhone's system. By extracting the event type from the file name, the researchers were able to label time-series acceleration data and video data, which will then serve as the ground truth in the model training and testing procedure.

During the testing sessions, the sensors (i.e., Apple Watch and iPhones) started recording before performing the activity and terminated data collection after finishing the entire trial sequence, ensuring that complete data of the activity were recorded. Because of such arrangement, the sensor-collected data comprise invalid signals or videos that happened in the following time intervals (Figure 4-17): 1) activation time of sensor to starting time of designed activity (i.e., $t_{\mathrm{s,\ sensor}}$ to $t_{\mathrm{s,\ activity}}$); 2) deactivation time of sensor to terminating time of designed activity (i.e., $t_{\mathrm{e,\ sensor}}$ to $t_{\mathrm{e,\ activity}}$). In order to label the data accurately, identifying and removing the invalid data patterns is necessary, which makes locating precise timestamps of starting and accomplishing activity crucial.

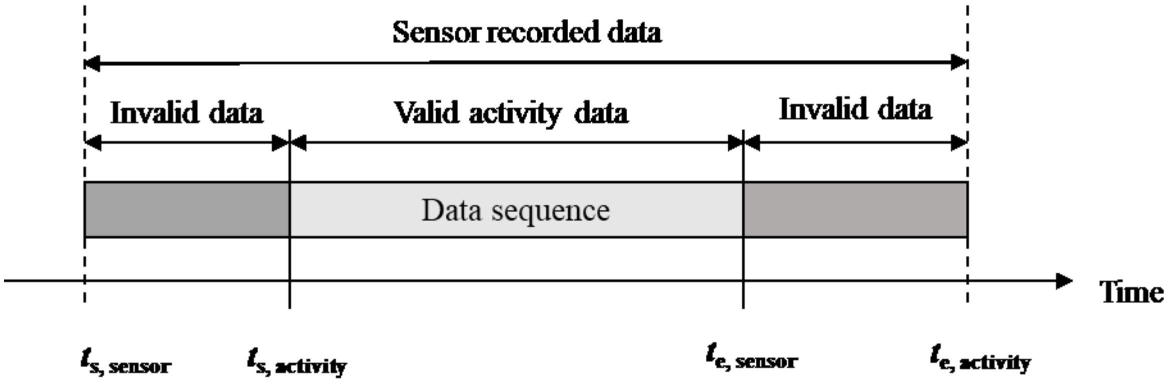


Figure 4-17 Illustration of invalid data in sensor-collected data

The required time information of video events could be directly obtained by reviewing raw videos. The manual observation is, however, time-consuming and not accurate due to inconsistent judgment from the reviewers. Though the timestamp judgment error in a single video clip is small (e.g., 10-millisecond level), the cumulated errors are significant, considering that the experimental activities were repeated continuously. In order to acquire accurate timestamps in an efficient manner, the researchers added additional actions right before performing the designed activity. Such warm-up action aims to introduce a strong signal before starting the activity, thereby guiding researchers to record the timestamps more efficiently and accurately. The current study required the participants to count down five seconds loudly and then conduct the activity immediately. By checking on the audio-time plot, the author could precisely locate the starting timestamp of the experimental activity. Furthermore, the author extracted the audio tracks from collected videos and plotted them in time. With the help of an audio analyzing algorithm, the precise starting time could be automatically located by searching the significant magnitude of the audio pulse at the beginning part.

Unlike videos presenting the experiment context, the acceleration data lacks attributes that allow the researchers to understand participants' states (i.e., whether the activity started or not). The remaining option is to visualize the time-series acceleration data and conduct a secondary analysis. For instance, Figure 4-18 shows the acceleration signals that refer to the "Hammering" performed by Participant No.1. An apparent magnitude of the signal pulse could be found after a silent period, leading to finding the exact starting time of "Hammering" (i.e., dash line in Figure 4-18) in the local clock. However, the acceleration signal vibrations are not significant, such as the "Lifting Brick" acceleration signal shown in Figure 4-19, resulting in considerable timestamp errors when determining the starting time. In order to overcome such limitation, the author also designed a warm-up action exclusive for acceleration data collection, which required the participants to wave the dominant hand (i.e., the hand that wore the Apple Watch) a couple of times. The strenuous movement of the wrist caused a cyclical and significant magnitude of signal pulse in the acceleration-time plot (Figure 4-19), which allows for precisely locating warming-up time in the internal clock system and gives a timestamped anchor to all the activity events in the acceleration data. Since cameras also record such warming-up actions, the offset of the time system between sensors could be calculated by subtracting the timestamp of the waving hand read in acceleration data and video data. Therefore, the common time system was established based on the calculated time offset. The timestamps were then transformed into the timestamp frame, thereby synchronizing the time from multiple sources of sensors.

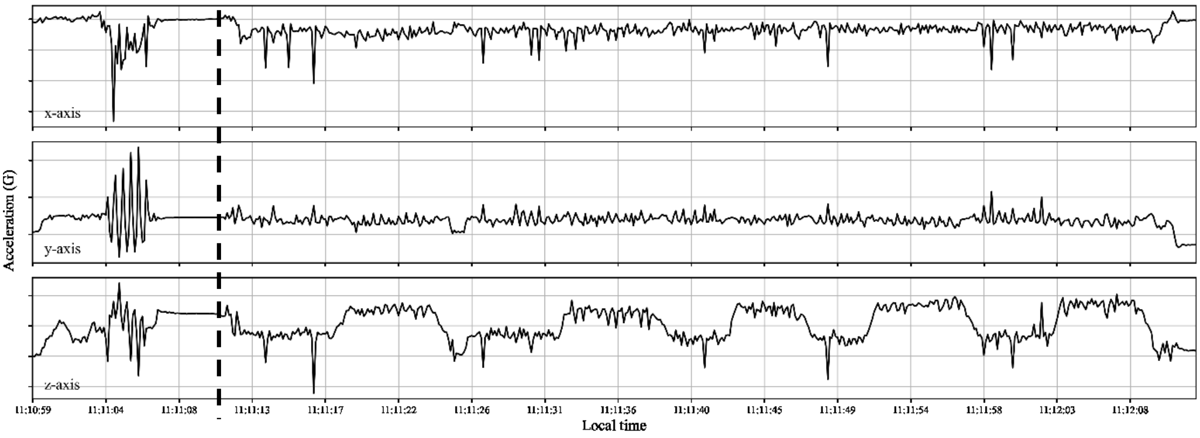


Figure 4-18 Example acceleration signals collected from "Hammering" activity

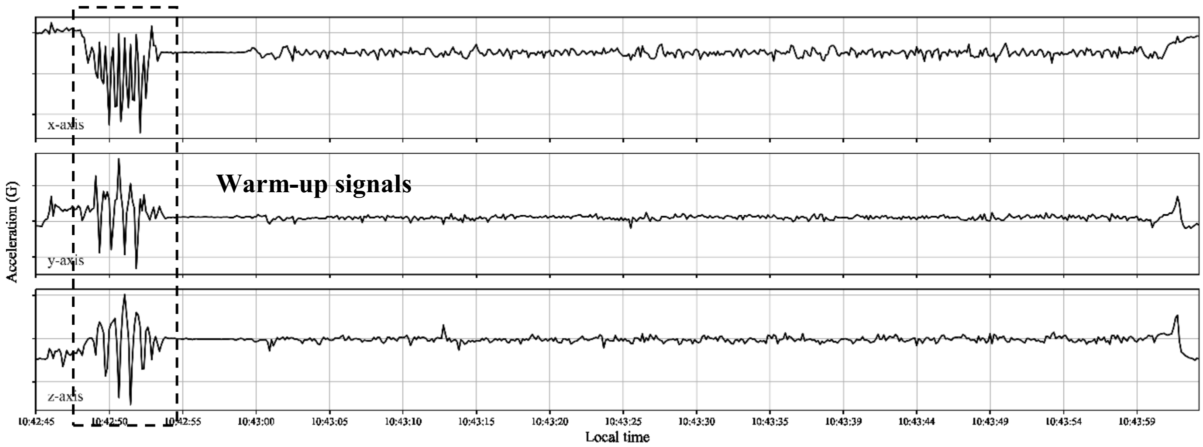


Figure 4-19 Example acceleration signal of "Lifting Brick"

### 4.6.3 Classifier training and evaluation

The current study tests various window lengths to achieve the best performance and, afterward, determines the optimal size based on model performance. The previous practice conducted by the author showed that the 1.5-second window size is an optimal setting for obtaining an accurate action recognition model (Gong et al., 2022). Due to the different data natures and experiment settings, the current study chose to test window length in a wide range rather than directly using the 1.5-second window. Therefore, the current study tested multiple window sizes that range from 0.5 seconds to 4.0 seconds with a step size of 0.5 seconds (i.e., 0.5, 1.0, 1.5, 2.0, 2.5, 3, 3.5, and 4.0 s). Considering the minimum window required for an activity task and the test experience, the window size was set into multiple values for further testing. The window sizes started from 0.5 seconds to 4.0 seconds with a strip of 0.5 seconds.

The data collected in this experiment is in the form of a time series, with each observation having a timestamp, thereby including temporal features. In such cases, the forward-chaining or expanding window cross-validation is a commonly used method for validating time-series models due to its consideration of the temporal dependence between observations. Specifically, this method employs earlier data for model training and later data for validation and testing.

However, this method is inappropriate when significant temporal dependence in the residuals or later observations significantly depends on earlier observations. Based on the repetitive temporal features of experiment data, this study recommends cross-validation for testing the effectiveness of proposed algorithms. In this method, the original data is divided into exclusive subsets of approximately equal size, with each subset being used once for validation and the other subsets used for model training. In this method, the original data is divided into exclusive subsets of equal size, with each subset being used once for validation and the other subsets used for model training. The proposed validation approach, aiming to minimize overfitting errors, divides the entire dataset into three subsets for training, validation, and testing purposes. In this study, one subset is reserved as the testing data, while the first three of the remaining four subsets are utilized for training, and the fourth subset is kept for validating the training results. Each data subset in the scenario can represent a full trial of the required activity performed during the experiment, ensuring equal size requirement and representation of the entire dataset. The consistency is due to participants being required to perform the same activity repeatedly for five trials, each lasting approximately the same duration. As a result, five models are generated, and associated validation error estimates are produced, which can be utilized to evaluate the performance of designed models and associated construction activity recognition framework. The evaluation process will be described in the following paragraphs, which include identifying the optimal size of the data pattern and evaluating the classifier's performance.

## 4.7 Result

In the current study, we tested the proposed framework using the data collected through laboratory experiments described above. The test's primary objective is to validate the effectiveness of decision-level fusion frameworks by comparing their classification performance to that of the single-sensor classification framework. The optimal window size is determined by analyzing the individual sensor's performance under different window sizes to achieve this. The identical window size is then adopted for conducting the fusion framework. The classification performance of the proposed decision-level fusion approach and the individual sensor's framework is evaluated using metrics such as classification accuracy, confusion matrix, precision, recall rate, and F1 score. By evaluating the fusion framework using three different fusion approaches, this study aims to recognize construction activities better and assess the efficacy of the proposed decision-level fusion framework.

### 4.7.1 Enhancing sensor-based action recognition: evaluating performance through optimized window size

The study evaluates the performance of two construction action recognition models based on individual sensor inputs, following the identical data processing, model training, and evaluation procedures illustrated in the previous section. Both tables present window sizes ranging from 0.5 to 4.0 seconds with a 0.5-second window increment and provide evaluation metrics such as average training accuracies, validation accuracies, testing accuracies, and standard deviation of testing accuracies. The optimal window size for the construction activity recognition model is determined through performance assessment using varying window sizes.

Table 4-3 shows the overall performance of the acceleration-based construction action recognition model with various window sizes. The model was trained, validated, and tested on segmented data, and the presented results demonstrate that the model achieved high average training accuracies across all window sizes, ranging from 98.73% to 99.29%. The highest average training accuracy of 99.29% was achieved with a window size of 4.0. Meanwhile, the model achieved the highest average testing accuracy of 73.64% with a 3.5-second window, followed by 73.12% with a 2.5-second window. The standard deviation of testing accuracy for the 3.5-second window is 0.73, which is lower than the 1.48 observed for the 2.5-second window. The average testing accuracy of 72.58% was observed for a 3.0-second window, which was the third-highest accuracy among all window sizes. However, the 3.0-second window setting generates a relatively high standard deviation of 1.34 for the testing accuracies. On the other hand, the model's average testing accuracy with a 1.0-second window size was 70.96%, which was relatively moderate compared to the other window sizes. Nonetheless, the 1.0-second window achieved the lowest standard deviation of 0.61, indicating its consistent performance and higher reliability across the dataset.

Table 4-4 for the video-based model. The current study employed segmented data for training, validation, and testing purposes in the acceleration-based model. Conversely, the video-based model used segmented signals only for testing. The mechanism can be found in the previous section, which elaborates on the data processing procedures. Therefore, the average training and validation accuracies for all window sizes are identical, which were found to be 92.33% and 76.69%, respectively, on average of five times evaluation. The table shows that the video-based construction action recognition model achieved the highest average testing accuracy of 71.24% with a window size of 3.5 seconds. The range of average testing accuracies is relatively small, with a difference of only 0.89% between the highest and lowest average testing values, which indicates that the model's performance was consistent across different window sizes. Additionally, the standard deviation variation for testing accuracies was relatively small, ranging from 3.03 to 3.62 for all window sizes. However, these standard deviation values were significant, suggesting moderate performance consistency when using different parts of the dataset.

Comparing Table 4-3 and Table 4-4 showed that the video-based model exhibited relatively low variation of average testing accuracies and their standard deviations compared to the acceleration-based model. Such a result indicates a more stable performance of the video-based model across varying window sizes and dataset selection. Therefore, the selection of window size for the video-based model did not significantly affect its performance. Conversely, a high variation of average testing accuracies and its standard deviation for the acceleration-based model was found in Table 4-3, revealing the high demand to optimize the window size. Hence, the window size optimization problem can be simplified as the optimization of window size for the acceleration model in the current study. Upon investigation of the results, it was found that a larger window size generally led to higher accuracy scores. Nonetheless, it is essential to consider the standard deviation of the testing accuracies to enhance the model's reliability across varying datasets. Furthermore, the utilization of a smaller window size can result in a more significant number of data points, which can lead to a better performance of the model by providing a representative sample of the population and mitigating the possible overfitting issue. After carefully considering the advantages and disadvantages of various window sizes, it is recommended to select the 1.0-second window as it presents a balanced and optimal choice for the study.

Table 4-3 Overall performance of acceleration-based construction action recognition model with different window sizes

| Window size (s) | Average training accuracies (%) | Average validation accuracies (%) | Average testing accuracies (%) | Test accuracies standard deviation (%) |
|---|---|---|---|---|
| 0.5 | 98.94 | 62.94 | 66.06 | 1.35 |
| 1.0 | 98.73 | 74.58 | 70.96 | 0.61 |
| 1.5 | 98.94 | 68.49 | 65.47 | 0.89 |
| 2.0 | 99.02 | 71.86 | 67.17 | 1.00 |
| 2.5 | 98.99 | 77.29 | 73.12 | 1.48 |
| 3.0 | 98.96 | 78.63 | 72.58 | 1.34 |
| 3.5 | 99.22 | 79.10 | 73.64 | 0.73 |
| 4.0 | 99.29 | 77.72 | 72.69 | 1.95 |

Table 4-4 Overall performance of video-based construction action recognition model with different window sizes

| Window size (s) | Average training accuracies (%) | Average validation accuracies (%) | Average testing accuracies (%) | Test accuracies standard deviation (%) |
|---|---|---|---|---|
| 0.5 | 92.33 | 76.69 | 70.35 | 3.13 |
| 1.0 | | | 71.49 | 3.03 |
| 1.5 | | | 70.53 | 3.16 |
| 2.0 | | | 70.78 | 3.42 |
| 2.5 | | | 70.89 | 3.18 |
| 3.0 | | | 70.84 | 3.32 |
| 3.5 | | | 71.24 | 3.04 |
| 4.0 | | | 71.12 | 3.62 |

### 4.7.2 Activity recognition results by using individual sensors' data

Despite the disparity in their data modalities and algorithm architectures, both models were subjected to the same dataset and scenario of training and testing, thereby providing a compelling comparative analysis method. In particular, a 1.0-second window size is advised as the optimal data segmentation choice for both models. In the meantime, a five-fold cross-validation approach was applied during the training process. The overall classification performance obtained from the individual sensors was presented in Table 4-5, including the average training, validation, and testing accuracies for accelerometer and video camera data sources. Upon reviewing the table, the accelerometer-based model showed higher overall performance with an average testing accuracy of 70.96% compared to the video camera-based model, which showed an average testing accuracy of 71.49%. Overall accuracy, mainly testing accuracy, is a widely used metric for evaluating the performance of classification models. At the same time, it has limitations in providing a detailed understanding of how the model performs for each class. In order to facilitate a deeper comprehension of the model's performance in classifying each activity, the current study examined the classification outcomes in the manner of confusion matrices. In this regard, Table 4-6 and Table 4-7 show the confusion matrices of the best-performing classifiers of an acceleration-based action recognition model and a video-based action recognition model, respectively. In the confusion matrixes, the actual category (columns) and predicted category (rows) are listed for each of the activities, including Traveling (TL), Lifting Brick (LB), Lifting Rebar (LR), Measuring Rebar (MR), Tying Rebar (TR), Hammering (HR), Drilling (DR), and Idling (ID). Each entry in the matrix corresponds to the intersection of a specific predicted activity with the actual activity, and its value indicates the number of instances in which the model made that specific prediction-actual pair. The matrix also includes additional classification metrics such as recall, precision, and the F1 score for each activity to evaluate the model's performance comprehensively. Specifically, high precision suggests that the model accurately predicts a category when it claims to do so, indicating fewer false positives (predicted instances incorrectly). Meanwhile, low precision means the model often mistakenly identifies an activity as a particular category, indicating a higher rate of false positives. Also, a high recall indicates that the model is good at identifying a particular category when it appears in the dataset, suggesting fewer false negatives, i.e., missed actual instances. In contrast, low recall implies that the model often fails to detect the category, indicating a high rate of false negatives. The F1 Score is taken into account as a balance metric of precision and recall. A high F1 score represents a robust model with a good balance between precision and recall. A low F1 score suggests that the model performs poorly in either precision, recall, or both, implying a need for improvement.

As presented in Table 4-6, the acceleration-based model demonstrates remarkable precision of above 88.0% in classifying "Traveling" (TL), "Lifting Brick" (LB), and "Lifting Rebar" (LR), which demonstrates superior performance in minimizing false positives for TL, LB, and LR. The precision accuracies for "Tying Rebar" (TR) and "Drilling" (DR) are lower than 60%, indicating space for refinement in classifying TR and DR. Particularly, the DR's precision is lower than 50%, meaning that more than half the time when the model predicted an activity as DR, it was incorrect. In the meantime, "Tying Rebar" (TR), "Lifting Brick" (LB), and "Hammering" (HR) were the top-performing categories with recall values of 88.8%, 87.5%, and 87.0%, respectively, which suggests a strong model performance in correctly identifying instances of TR, LB, and HR. Conversely, "Idling" (ID) fared the worst in the recall metric, with 16.5%, indicating that the model struggled to identify ID instances correctly. The F1 Score, which balances precision and recall, reaches high values for "Lifting Rebar" (LB) and "Hammering" (HR) at 0.88 and 0.85, respectively, suggesting overall optimal performance for these activity categories by the acceleration-based model.

Similarly, the confusion matrix in Table 4-7 represents the five-fold cross-validation results from the best-performed video-based action recognition model (300 epoch training with a 1.0-second window). The model exhibits exceptional precision in recognizing "Lifting Brick" (LB), "Hammering" (HR), and "Drilling" (DR), all of which show a precision exceeding 90%. The results emphasize the model's strength in classifying such activity categories with a low ratio of false positives. However, "Idling" (ID) and "Lifting Rebar" (LR) exhibit lower precision, below 60%, implying that these categories need model improvements for precise classification. In terms of recall, "Drilling" (DR) is the leading category with a recall of 94.9%, indicating the model's substantial success in accurately identifying instances of DR. TL, LB, LR, and TR also show impressive recall values of the above 80%, meaning the model's robustness in correctly identifying instances of these activities. Conversely, "Hammering" (HR) has the lowest recall value, at 60.5%, signifying that the model faces challenges in correctly identifying HR instances. The F1 Score suggests satisfactory performance for "Drilling" (DR), with an F1 Score of 0.97. So that the model optimally recognizes the "Drilling" (DR) activity.

Nevertheless, the F1 Score for “Measuring Rebar” (MR) is 0.62, indicating a relatively low overall classification performance that could be improved for this category.

Table 4-5 Performance of individual sensor-based construction action recognition models using 1.0-second windows

| Sensor source | Average training accuracy (%) | Average validation accuracies (%) | Average testing accuracies (%) |
|---|---|---|---|
| Accelerometer | 98.73 | 74.58 | 70.96 |
| Video camera | 92.33 | 76.69 | 71.49 |

Table 4-6 Confusion matrix of acceleration-based action recognition result (1.0-second window)

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 96 | 3 | 0 | 18 | 21 | 0 | 6 | 0 | 66.7 |
| | LB | 3 | 168 | 3 | 15 | 3 | 0 | 0 | 0 | 87.5 |
| | LR | 0 | 12 | 147 | 9 | 30 | 0 | 6 | 0 | 72.1 |
| | MR | 0 | 0 | 3 | 187 | 26 | 0 | 3 | 4 | 83.9 |
| | TR | 2 | 2 | 4 | 14 | 198 | 1 | 1 | 1 | 88.8 |
| | HR | 0 | 0 | 0 | 18 | 0 | 141 | 0 | 3 | 87.0 |
| | DR | 0 | 0 | 6 | 21 | 9 | 0 | 90 | 12 | 65.2 |
| | ID | 6 | 6 | 3 | 15 | 49 | 27 | 76 | 36 | 16.5 |
| | Precision (%) | 89.7 | 88.0 | 88.6 | 63.0 | 58.9 | 83.4 | 49.5 | 64.3 | |
| | F1 Score | 0.76 | 0.88 | 0.79 | 0.72 | 0.71 | 0.85 | 0.56 | 0.65 | |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

Table 4-7 Confusion matrix of video-based action recognition result (1.0-second window, 300 epoch training)

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 116 | 13 | 12 | 0 | 3 | 0 | 0 | 0 | 80.6 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| LB | 6 | 167 | 14 | 0 | 5 | 0 | 0 | 0 | 87.0 |
| LR | 11 | 0 | 170 | 15 | 8 | 0 | 0 | 0 | 83.3 |
| MR | 5 | 0 | 5 | 125 | 88 | 0 | 0 | 0 | 56.1 |
| TR | 2 | 0 | 1 | 40 | 180 | 0 | 0 | 0 | 80.7 |
| HR | 0 | 0 | 0 | 0 | 0 | 98 | 0 | 64 | 60.5 |
| DR | 0 | 0 | 0 | 0 | 0 | 0 | 131 | 7 | 94.9 |
| ID | 33 | 0 | 96 | 0 | 0 | 0 | 0 | 89 | 40.8 |
| Precision (%) | 67.1 | 92.8 | 57.0 | 69.4 | 63.4 | 100.0 | 100.0 | 55.6 | |
| F1 Score | 0.73 | 0.90 | 0.68 | 0.62 | 0.71 | 0.75 | 0.97 | 0.70 | |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

### 4.7.3 Activity recognition results by using the decision-level fusion method

Table 4-8 presents a detailed summary of the average testing accuracies of construction activity recognition models, which utilize a decision-level fusion of results derived from multiple sensor sources. For comparative analysis, the table also lists the classification results generated by employing a single sensor source. As shown in the table, the acceleration-based model yielded an average testing accuracy of 70.96%, and the video-based classifier delivered an average accuracy of approximately 71.49%.

As for the decision-level fusion methods, the present research incorporates the Dempster-Shafer method (DS) as the foundational model and then expands to include variations such as the Weighted Dempster-Shafer method (WDS), the Topk Weighted Dempster-Shafer method (TopkWDS), and the Thresholding Weighted Dempster-Shafer method (TWDS). Multiple weighing alternatives are available in each of these weighted methodologies, encompassing precision, recall, and F1 score obtained from the validation process, depending upon the specific demand of research. It should be noted that for the TopkWDS and TWDS, the parameters for weights are also included in Table 4-8. The conventional Dempster-Shafer method (DS) notably increased testing accuracy to 80.43%, outperforming individual sensor processing approaches. The performance was relatively similar when using the Weighted Dempster-Shafer method (WDS). The accuracies reached 80.36%, 80.24%, and 80.46%, respectively, when setting the precision, recall, and F1 scores as weights. The Topk Weighted Dempster-Shafer method (TopkWDS) showed further improvements. This method assigned weights based on the top-k values, where k was defined separately for acceleration ($k_{\text{acc}}$) and video ($k_{\text{vd}}$) data. For $k_{\text{acc}} = 2$ and $k_{\text{vd}} = 7$, this method achieved testing accuracies over 83%. In particular, when using recall accuracy as the weight, the testing accuracy achieved the highest accuracy of 83.67%. According to the table, the Thresholding Weighted Dempster-Shafer method (TWDS) was the most effective in the current experiment. The method used weights derived from threshold values set for acceleration ($e_{\text{acc}}$) and video ($e_{\text{vd}}$) data. With parameter settings of $e_{\text{acc}} = 0.87$ and $e_{\text{vd}} = 1.00$, the TWDS method led to a testing accuracy of 85.39% when adopting the precision value as the weight towards each activity category.

Table 4-8 Overall performance of construction action recognition model using different methods

| Method | Testing accuracies (%) |
|---|---|
| **Individual sensor processing**<br>(Sensor source – algorithm architecture) | |
| Acceleration-BiLSTM | 70.96 |
| Video-ResNets | 71.49 |
| **Decision-level fusion of multi-sensor results** | |
| **D**empster-**S**hafer method (**DS**) | 80.43 |
| **W**eighted **D**empster-Shafer method (**WDS**) | |
| weight = *Precision* | 80.36 |
| weight = *Recall* | 80.24 |
| weight = *F1* | **80.46** |
| **Topk W**eighted **D**empster-Shafer method (**TopkWDS**) | |
| weight = *Precision* ($k_{\text{acc}} = 2$, $k_{\text{vd}} = 7$) | 83.36 |
| weight = *Recall* ($k_{\text{acc}} = 2$, $k_{\text{vd}} = 7$) | **83.67** |

| | |
|---|---|
| weight = *F1* ($k_{\text{acc}} = 2$, $k_{\text{vd}} = 7$) | 83.62 |
| **T**hresholding **W**eighted **D**empster-**S**hafer method (**TWDS**) | |
| weight = *Precision* ($e_{\text{acc}} = 0.87$, $e_{\text{vd}} = 1.00$) | **85.39** |
| weight = *Recall* ($e_{\text{acc}} = 0.87$, $e_{\text{vd}} = 1.00$) | 85.32 |
| weight = *F1* ($e_{\text{acc}} = 0.87$, $e_{\text{vd}} = 1.00$) | 85.37 |

To better illustrate the performance of each decision-level fusion method mentioned in Table 4-8, representative examples of corresponding confusion matrices can be found in Table 4-9 through Table 4-12. The parameter selection is the same as that shown in Table 4-8.

Table 4-9 displays the performance of the Dempster-Shafer (DS) method for action recognition. The method exhibits impressive precision levels for activities such as “Traveling” (93.3%), “Lifting Brick” (92.1%), and “Lifting Rebar” (91.2%), demonstrating its effectiveness in reducing false positives for classifying the mentioned activities. However, the precision accuracy for “Drilling” (55.3%) falls below 60%, indicating a need for improvement in this category. On the recall front, “Hammering” performs the best with 95.1%, followed by “Lifting Brick” at 91.1%, and “Tying Rebar” at 87.5%, showing the method’s robust capability in correctly identifying instances of these activities. Conversely, “Idling” exhibits poor performance in the recall metric, scoring just 36.2%, signifying the DS method’s difficulty in correctly identifying ID instances. The F1 Score, a measure balancing precision and recall, attains high values for “Lifting Brick” (0.92), “Hammering” (0.92), and “Traveling” (0.90), suggesting an overall robust performance of the DS method for these activity categories.

Table 4-10 indicates the performance of the F1-Weighted Dempster-Shafer (F1-WDS) method, using F1 scores for validation as weights for each activity category in the DS combination. The table shows that the WDS method shows modest improvement compared to the baseline Dempster-Shafer (DS) model. Similar to the DS method, the WDS method demonstrates reliable performance in mitigating false positives for activities such as “Traveling” (93.3%), “Lifting Brick” (92.1%), and “Lifting Rebar” (91.2%). On the recall spectrum, “Hammering” excels with a score of 95.1%. However, "Idling" performs poorly, with a recall of 35.8%. The F1 Score presents high values for "Lifting Brick," "Hammering," and "Traveling," indicating a balanced performance for these activities.

Table 4-11 displays a confusion matrix for the decision-level fusion action recognition model, which employs the Topk Weighted Dempster-Shafer (TopkWDS) method. In the given example, the weights assigned during the fusion process were derived from the recall values calculated from the validation result. The TopkWDS model demonstrates commendable precision levels exceeding 90% for activities such as "Traveling" (92.8%), "Lifting Brick" (94.8), and "Hammering" (91.9%). Nevertheless, the precision for "Drilling" is notably lower at 61.3%, suggesting a need for enhancement in this area. When viewed from recall values, "Lifting Brick" (94.3%) and "Hammering" (91.4%) emerge as the highest-performing categories, both exceeding 90%. These high precision scores highlight the model's proficiency in accurately classifying instances of "Traveling" (TL), "Lifting Brick" (LB), and "Hammering" (HR). However, "Idling" (ID) has the lowest recall ratio at 50%, indicating that the model misclassified half of the actual ID instances, suggesting difficulties in identifying this category. In terms of the F1 Score, "Lifting Brick" (LB) is notable, with an impressive score of 0.95, indicating high accuracy and efficiency in identifying LB instances. Furthermore, both "Traveling" (TL) and "Drilling" (DR) also exhibit excellent performance, as evidenced by their F1 scores surpassing 0.90, indicating a solid balance between precision and recall in classifying these activities. In contrast, the F1 Score for DR is at 0.70, the lowest among all categories in the given model, underscoring a need for improvement in this area.

Table 4-12 offers insight into a decision-level fusion model based on the Thresholding Weighted Dempster-Shafer (TWDS) technique, focusing primarily on identifying construction activities. The weight used in this example is the precision value of each category in the validation process. Notably, the model's precision is particularly prominent in activities such as “Lifting Brick” (LB), “Hammering” (HR), and “Traveling” (TL), with corresponding precision values of 94.4%, 93.5%, and 89.6%. Notably, the precision measurements for identifying “Lifting Rebar” and “Drilling” in construction practices are relatively low, with a score of 74.3% and 78.6%, respectively. A survey of the recall values reveals “Lifting Brick” (LB) at an impressive 95.8%, closely followed by “Traveling” (89.6%) and “Lifting Rebar” (89.2%), further underscoring the TWDS technique’s effectiveness. However, a performance downturn corresponding to the “Idling” (ID) category is found, which executes a low recall value of 53.7%, pointing to potential identification challenges applying the TWDS framework in construction activity recognition. Lastly, the F1 Score assigns high scores to LB (0.95), TL (0.90), and HR (0.91), confirming the robustness of the model in these categories by effectively minimizing both false positives and false negatives in predictions. Conversely, LR, DR, and ID show weak classification performance with potential for optimization. The classification of LR scored 0.81, while DR and ID scored slightly better, with 0.83 and 0.84, respectively.

Table 4-9 Confusion matrix of decision-level fusion action recognition using the Dempster-Shafer (DS) method

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 126 | 2 | 0 | 6 | 4 | 0 | 6 | 0 | 87.5 |
| | LB | 2 | 175 | 8 | 7 | 0 | 0 | 0 | 0 | 91.1 |
| | LR | 1 | 9 | 165 | 6 | 16 | 1 | 6 | 0 | 80.9 |
| | MR | 0 | 0 | 0 | 193 | 28 | 0 | 0 | 2 | 86.5 |
| | TR | 0 | 0 | 3 | 16 | 203 | 0 | 1 | 0 | 91.0 |
| | HR | 0 | 0 | 0 | 5 | 0 | 154 | 0 | 3 | 95.1 |
| | DR | 0 | 0 | 3 | 6 | 1 | 0 | 110 | 18 | 79.7 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ID | 6 | 4 | 2 | 8 | 26 | 17 | 76 | 79 | 36.2 |
| | Precision (%) | 93.3 | 92.1 | 91.2 | 78.1 | 73.0 | 89.5 | 55.3 | 77.5 | |
| | F1 Score | 0.90 | 0.92 | 0.86 | 0.82 | 0.81 | 0.92 | 0.65 | 0.79 | |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

Table 4-10 Confusion matrix of decision-level fusion action recognition using F1 score-Weighted Dempster–Shafer method (F1-WDS)

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 126 | 2 | 0 | 6 | 4 | 0 | 6 | 0 | 87.5 |
| | LB | 2 | 175 | 8 | 7 | 0 | 0 | 0 | 0 | 91.1 |
| | LR | 1 | 9 | 165 | 6 | 16 | 1 | 6 | 0 | 80.9 |
| | MR | 0 | 0 | 0 | 193 | 29 | 0 | 0 | 1 | 86.5 |
| | TR | 0 | 0 | 3 | 16 | 203 | 0 | 1 | 0 | 91.0 |
| | HR | 0 | 0 | 0 | 5 | 0 | 154 | 0 | 3 | 95.1 |
| | DR | 0 | 0 | 3 | 6 | 1 | 0 | 110 | 18 | 79.7 |
| | ID | 6 | 4 | 2 | 8 | 26 | 18 | 76 | 78 | 35.8 |
| | Precision (%) | 93.3 | 92.1 | 91.2 | 78.1 | 72.8 | 89.0 | 55.3 | 78.0 | |
| | F1 Score | 0.90 | 0.92 | 0.86 | 0.82 | 0.81 | 0.92 | 0.65 | 0.79 | |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

Table 4-11 Confusion matrix of decision-level fusion action recognition using Recall-Topk Weighted Dempster-Shafer method (Recall-TopkWDS)

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 128 | 1 | 3 | 4 | 4 | 0 | 4 | 0 | 88.9 |
| | LB | 2 | 181 | 4 | 5 | 0 | 0 | 0 | 0 | 94.3 |

|  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|
|  | LR | 1 | 8 | 176 | 4 | 10 | 0 | 5 | 0 | 86.3 |
|  | MR | 0 | 0 | 1 | 195 | 26 | 0 | 0 | 1 | 87.4 |
|  | TR | 0 | 0 | 3 | 16 | 203 | 0 | 1 | 0 | 91.0 |
|  | HR | 0 | 0 | 0 | 5 | 0 | 148 | 0 | 9 | 91.4 |
|  | DR | 0 | 0 | 3 | 4 | 0 | 0 | 114 | 17 | 82.6 |
|  | ID | 7 | 1 | 21 | 3 | 2 | 13 | 62 | 109 | 50.0 |
|  | Precision (%) | 92.8 | 94.8 | 83.4 | 82.6 | 82.9 | 91.9 | 61.3 | 80.1 |  |
|  | F1 Score | 0.91 | 0.95 | 0.85 | 0.85 | 0.87 | 0.92 | 0.70 | 0.81 |  |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

Table 4-12 Confusion matrix of decision-level fusion action recognition using Precision-Thresholding Weighted Dempster-Shafer method (Precision-TWDS)

| Activity | True category | TL | LB | LR | MR | TR | HR | DR | ID | Recall (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Predicted category | TL | 129 | 2 | 5 | 2 | 3 | 0 | 3 | 0 | 89.6 |
|  | LB | 2 | 184 | 1 | 5 | 0 | 0 | 0 | 0 | 95.8 |
|  | LR | 1 | 8 | 182 | 1 | 9 | 0 | 3 | 0 | 89.2 |
|  | MR | 1 | 0 | 1 | 193 | 27 | 0 | 0 | 1 | 86.5 |
|  | TR | 0 | 0 | 2 | 12 | 208 | 0 | 1 | 0 | 93.3 |
|  | HR | 0 | 0 | 0 | 4 | 0 | 144 | 0 | 14 | 88.9 |
|  | DR | 0 | 0 | 3 | 2 | 0 | 0 | 121 | 12 | 87.7 |
|  | ID | 11 | 1 | 51 | 2 | 0 | 10 | 26 | 117 | 53.7 |
|  | Precision (%) | 89.6 | 94.4 | 74.3 | 87.3 | 84.2 | 93.5 | 78.6 | 81.3 |  |
|  | F1 Score | 0.90 | 0.95 | 0.81 | 0.87 | 0.89 | 0.91 | 0.83 | 0.84 |  |

* Traveling: TL, Lifting Brick: LB, Lifting Rebar: LR, Measuring Rebar: MR, Tying Rebar: TR, Hammering: HR, Drilling: DR, Idling: ID

## 4.8 Discussions

In the current study, four decision-level fusion methodologies were developed in construction action recognition, including the Dempster-Shafer method (DS), the Weighted Dempster-Shafer method (WDS), the Topk Weighted Dempster-Shafer method (TopkWDS), and the Thresholding Weighted Dempster-Shafer method (TWDS). In order to validate the fusion methods, this study conducted an experiment that simulated eight construction site activities. Acceleration and videos were collected simultaneously, and learning algorithms were used to train action recognition models using single sensor sources, resulting in one acceleration-based and one video-based model. The preliminary results obtained from these two models were then input into the proposed decision-level fusion framework, and the effectiveness of the fusion methods was evaluated by comparing the testing accuracy and confusion matrix.

### 4.8.1 Decision-level fusion validation

The acceleration-based model demonstrated an average testing accuracy of 70.96%, while the video-based model slightly outperformed it with an average testing accuracy of 71.49%. In addition, the decision-level fusion methods, namely DS, WDS, TopkWDS, and TWDS, exhibited superior performance compared to the individual sensor-based models, validating the effectiveness of the fusion strategy. The DS method notably enhanced testing accuracy to 80.43%, while the WDS method achieved accuracies of 80.36%, 80.24%, and 80.46% when precision, recall, and F1 scores were set as weights, respectively. The TopkWDS method further improved upon these results, achieving testing accuracies over 83%, with the highest accuracy reaching 83.67% when using recall accuracy as the weight (Recall-TopkWDS). However, the most effective method in this study was the TWDS method, which achieved the highest testing accuracy of 85.67%. In the meantime, the DS method demonstrated high precision for “Traveling” (93.3%), “Lifting Brick” (92.1%), and “Lifting Rebar” (91.2%) but struggled with “Drilling” (55.3%) and “Idling” (36.2% recall). The WDS method maintained similar precision levels and slightly improved “Idling” recall (35.8%). TopkWDS exceeded 90% precision for “Traveling” (92.8%), “Lifting Brick” (94.8), and “Hammering” (91.9%) but fell short in “Drilling” (61.3%) and “Idling” (50% recall). TWDS showed high precision for “Lifting Brick” (94.4%), “Hammering” (93.5%), and “Traveling” (89.6%) but faced challenges in “Lifting Rebar” (74.3%) and “Drilling” (78.6%), with a low recall for “Idling” (53.7%). In terms of F1 scores, DS, F1-WDS, and TopkWDS, the models achieved high values for “Lifting Brick” (0.92, 0.92, 0.95 respectively), “Hammering” (0.92, 0.92, 0.91), and “Traveling” (0.90, 0.90, 0.90), indicating a balanced performance in classifying these activities. However, “Drilling” consistently scored lower, with the lowest score observed in TopkWDS (0.70).

Upon comparative analysis, it is evident that all four decision-level fusion methods, DS, F1-WDS, TopkWDS, and TWDS, outperform the model trained solely on single-sensor data. Specifically, the WDS method exhibits a modest enhancement in performance relative to the DS method. Meanwhile, the introduction of weight filtering processes in the TopkWDS and TWDS methods yields substantial augmentations in overall testing accuracies when compared with the WDS method. The system’s superior performance can be attributed to the effective implementation of the fusion method at the decision level, which leverages the unique strengths inherent in both acceleration and video data, thereby outperforming the single-sensor method. These findings validate the effectiveness of decision-level fusion methods but also highlight the need for further optimization to enhance poor performance activity types, such as “Iding.”

### 4.8.2 Sensor credibility and weight selection

The preceding section validates the effectiveness of the decision-level fusion method, suggesting that this approach could serve as a feasible alternative to enhance the performance of the entire sensor system by overcoming the variability inherent in individual classifiers. However, the modest improvement observed with the DS method over the single-sensor model indicates that further optimization is required. This leads to considering another critical factor in sensor system development, which is the unequal trust levels associated with different sensors.

The confusion matrices of the acceleration-based and video-based models reveal distinct performance characteristics (e.g., precision, recall, and F1 score) for each activity category despite similar overall testing accuracy levels (around 70%). This discrepancy suggests that the credibility or trust level of a sensor may vary for different activity events. Generally, each sensor type has its strengths and weaknesses in detecting specific activity patterns. For instance, as shown in Table 4-6, the acceleration model exhibits poor precision for “Drilling” (49.5%), while the video model achieves 100% precision for the same activity using identical testing data. The contrast can be attributed to the inherent nature of the data. In specific, the “Drilling” activity in the experiment, which involves holding a drilling machine to drill a wood slab, includes minimal wrist movement. Such a pattern is detectable in the acceleration signal but can also be found in other activities involving minor wrist vibrations, such as “Idling,” leading to significant misclassification errors. Conversely, the video model primarily identifies activities based on image information. The “Drilling” activity in this experiment has a distinct characteristic - the presence of a drilling machine. If the model detects the drilling machine in the test video, it is highly likely to classify the activity as “Drilling.” This example underscores the significance of considering each sensor type’s credibility when developing a sensor system.

The current study introduces a weighted method using the Dempster-Shafer approach, termed WDS, to address the errors introduced by the unequal trust levels of sensors. However, the marginal improvement observed between WDS and DS underscores the limitations of directly applying weights to balance unequal sensor trust, suggesting that the direct application of weights may need to adequately account for the inherent variability and uncertainty in sensor data. In response to this challenge, the study proposes a weight filtering process prior to the assignment of weights in the combination of sensor estimates. The rationale for this approach stems from the mechanism of the DS combination, which considers all uncertainties, including extremely small values representing improbable events in possibility theory. These small values are factored into the combination calculation, thereby undermining the influence of high-probability predictions and amplifying the influence of low-probability predictions. To mitigate this issue, two methods, TopK and Thresholding, are proposed to filter out extremely small values in the single prediction from each sensor. Specifically, TopK involves selecting the largest k predictions, while Thresholding involves selecting predictions with a possibility greater than a predetermined value. The unselected portion is combined into a single instance to maintain the integrity of the DS method. The introduction of weight filtering processes in the TopkWDS and TWDS methods results in substantial improvements in overall testing accuracies compared to the WDS method. This finding demonstrates the essential role of weight filtering in balancing sensor credibility and emphasizes the potential of such processes to enhance model performance.

In the current study, we explored using precision, recall, and F1 score metrics calculated from the validation results during training. The rationale for using metrics derived from the validation set, as opposed to those from the testing or training results, is to prevent exposure of testing data set information. This exploration of different metrics aims to understand the differential impacts of using various weight types. The results reveal minimal differences when employing different metrics. Despite these minor variations, the choice of metric as a weight can still be important based on specific needs. For instance, in safety-related topics such as identifying hazardous activities during construction, minimizing false negatives (missed alarms) is more critical than reducing false positives (false alarms), which implies a higher priority for recall in safety cases. Therefore, the selection of metrics, serving as an alternative weight, can be customized based on specific requirements in future applications, thereby enhancing the adaptability and effectiveness of the model.

### 4.8.3 Limitations

Despite implementing decision-level fusion and introducing filtered weights to balance both model bias and sensor credibility biases, the prediction performance for specific categories, such as "Idling," remains poor. The poor performance is mainly due to the fundamental limitations in the base predictions from the acceleration and video models. While the proposed fusion method has slightly improved performance, the limitations can be primarily attributed to data constraints and base model restrictions. Despite conducting this study in a controlled lab environment, the study tried to simulate conditions as close as possible to a real construction site. For instance, the background was intentionally made disorderly, which posed initial classification challenges. As for model limitations, we sought to leverage the primary advantage of decision-level fusion, i.e., integrating multiple sources without spending excessive computational costs. Consequently, the study employed basic models in the activity recognition deep learning algorithms. Another potential limitation lies in fully exploiting the benefits of decision-level fusion. As discussed in the previous chapter, decision-level fusion not only enhances computational efficiency but also provides substantial flexibility for constructing large sensor fusion networks. Such flexibility implies a high level of resilience, ensuring the availability of the final output even if several sensor nodes malfunction. However, current research has yet to validate the resilience aspect of decision-level fusion, highlighting an area for future exploration.

### 4.8.4 Future research

Future studies will first focus on collecting more data from laboratory environments to overcome the limitations caused by the dataset's limited size, thereby ensuring that models generalize more effectively, minimize overfitting, and mirror real-world patterns more accurately. The current study has recruited three additional participants (excluding the existing participants) to perform the same activities designed under the identical experimental setup, utilizing smartphones to record videos from three perspectives and smartwatches to collect acceleration data simultaneously. This expanded test, conducted in a new laboratory environment, introduces background variations in the videos, aiming to achieve a more generalized model. Additionally, the feasibility of the proposed fusion framework will be assessed in actual construction sites using video and acceleration data from workers on ongoing projects.

The current study utilizes data from three participants, recorded from three different points of view. It adopted a five-fold cross-validation strategy to train and validate the model, where the satisfactory testing accuracies underscore the model's generalizability across video recording angles. However, generalizability across participants has not been validated. Future research will employ a Leave-One-Subject-Out Validation (LOSOV) approach with the expanded dataset to overcome this limitation. Specifically, in each iteration, one participant's data will be designated as the testing data, another as the validation data, and the remaining four participants' data will serve as the training dataset. This process, repeated six times with unique combinations of training, testing, and validation datasets, aims to address variances from individual participants, recording angles, and backgrounds. Should the model achieve acceptable average testing accuracy across these iterations, as per the LOSOV framework, it will affirm its capability to generalize across different individuals, angles, and backgrounds, bolstering sensor fusion methods' effectiveness.

Considering the dynamic environment of construction sites, a comprehensive sensor network employing various types of sensors is essential. Given that sensors on construction sites are prone to damage, this study opts not to pursue high-performance algorithms that rely on single sensor data—these often entail complex structures, high computational costs, and lower resilience. Instead, it aims to leverage the full potential of diverse sensor resources through fusion methods to achieve optimal monitoring results. This study employed a Dempster-Shafer theory-based approach as the baseline model, with plans to evaluate additional models beyond the DS method. Furthermore, future research will investigate fusion performance at data-level, feature-level, and hybrid-level levels to develop an optimal sensor network resistant to noise and environmental variations. Furthermore, the current study utilized the validation accuracy of each category from different models to represent the credibility of a specific data source for a particular activity. In future research, additional methods for measuring the uncertainty of the same event from different resources will be explored, including the entropy method. In addition to classification accuracy, future validation will assess the fusion network's stability, resilience, and flexibility. A potential approach involves using the Monte Carlo method to randomly omit specific data points, enabling a comparison of decision-level fusion methods' resilience with and without missing data, thereby demonstrating their robustness.

## 4.9 Conclusions

This study developed and evaluated four decision-level fusion methodologies for construction action recognition: the Dempster-Shafer method (DS), the Weighted Dempster-Shafer method (WDS), the Topk Weighted Dempster-Shafer method (TopkWDS), and the Thresholding Weighted Dempster-Shafer method (TWDS). These methods were tested using data from an experiment simulating eight construction site activities, with acceleration and video data collected simultaneously. The results demonstrated that decision-level fusion methods outperformed the single-sensor models, validating the effectiveness of the fusion strategy. However, the performance of specific activity types, such as "Idling," remained suboptimal, indicating the need for further optimization.

The study also explored the impact of sensor credibility and weight selection on the performance of the fusion methods. The results highlighted the limitations of directly applying weights to balance unequal sensor trust and proposed a weight-filtering process to address this issue. This process, implemented in the TopkWDS and TWDS methods, led to substantial improvements in overall testing accuracies, underscoring the potential of such processes to enhance model performance.

Despite these advancements, the prediction performance for specific categories could have improved due to fundamental limitations in the base predictions from the acceleration and video models. Future research will aim to address these limitations by collecting more data, preferably from real construction sites, and segmenting larger window sizes. Additionally, future studies will seek to validate the flexibility of decision-level-based sensor fusion networks and explore other fusion methods, including data-level, feature-level, and hybrid-level fusion methods. These efforts will contribute to developing an optimal sensor network that is highly resistant to noise and environmental variations, thereby enhancing the robustness and applicability of sensor fusion methods in real-world scenarios.

# CHAPTER 5 FIELD VALIDATION OF BEACON-BASED INDOOR TRACKING AND LOCALIZATION SYSTEM FOR CONSTRUCTION WORKERS[3]

## 5.1 Background

Since construction operations require various construction entities (e.g., workers, equipment, and materials) to engage, tracking and locating these entities are necessary for multiple applications in construction sites, such as safety management (Cai & Cai, 2020; Khoury & Kamat, 2009), resource optimization (Dzeng et al., 2014), and progress monitoring (Cheng et al., 2013). For example, knowing the geographic positions of workers, equipment, and hazardous zones allows us to identify and analyze the proximity of these entities. The workers will be alarmed when excessive proximity is identified during the operation, and therefore, the potential near-miss accident could be prevented (Liu et al., 2018; Papaioannou et al., 2016). Considering the large number of related entities and the large scale of the construction site, traditional localization with manual observation is labor-intensive and error-prone (Zhang et al., 2013), which makes the automated approach an essential role in tracking construction entities on the site.

Among various wireless technologies (e.g., Radio Frequency Identification (RFID), Global Positioning Systems (GPS), and Ultra-Wideband (UWB)) for tracking and locating construction entities, Bluetooth Low Energy (BLE) beacons have comparative advantages of 1) a low amount of infrastructure setting (Zhao et al., 2019), 2) flexible installation (Urano et

[3] This chapter is based on the manuscript titled “Field Validation of Beacon-Based Indoor Tracking and Localization System for Construction Workers” which has been submitted to the KSCE Journal of Civil Engineering and is currently under review, with the thesis author serving as the first author of the paper.

al., 2017), 3) accessible to scalable both indoor and outdoor (Ng et al., 2020), and 4) cost-effective (Park et al., 2017). For example, unlike UWB, which requires a continuous power supply, a battery-less BLE beacon is more flexible to deploy in a fast-changing environment (e.g., a construction site) (Khoury & Kamat, 2009). In contrast to wireless technologies, such as RFID and magnetic field, that need time-consuming calibration, the BLE beacon is capable of calibrating easily, therefore minimizing the infrastructure requirements (Park, Marks, et al., 2016). Due to these advantages and unique features of BLE beacons, previous studies have applied beacon-based tracking to construction workers (Park et al., 2017), resources (Shen et al., 2008), and vehicles (Lu et al., 2007).

Despite the advantages of BLE beacons, one of the critical issues is the accuracy and reliability of beacon-based tracking and localization. The typical scenario of beacon-based tracking and localization is based on the distance measure (i.e., Rx power level approach) between the beacon and the receiver (e.g., smartphone) by using the characteristics of beacon signals that the signal strength would gradually decrease during propagation (Subhan et al., 2011). By using the estimated distances from multiple beacons (at least three), the receiver's position can be detected through trilateration methods (Elnahrawy et al., 2004; Han et al., 2007). However, the distance estimation is not always stable because the received signal tends to fluctuate as it is affected by environmental factors such as temperature and humidity (Guidara et al., 2018). Also, the designed bandwidth of BLE technology does not allow the signal to penetrate obstacles like walls. Therefore, the signal received is the combination of signals from multiple paths, including directly received signals or signals reflected by walls, which could lead to inaccurate distance estimation (Faragher & Harle, 2014). This issue would be more significant, especially when deploying multiple beacons in the same area (Mackey et al., 2018). To mitigate signal frustration, previous studies have proposed and tested mathematical approaches to filter out noisy signals, such as the Bayes filter, Kalman Filter (KF), Extended Kalman Filter (EKF), and Particle Filter (PF) (Xu et al., 2021). However, most studies mainly focused on signal noise and fluctuation during signal propagation without fully considering the impact of diverse environmental conditions on beacon signals.

In this regard, the influence of diverse environmental conditions and various application scenarios on signal strength at construction sites has been examined through multiple field tests. Beacon signals were collected from widely employed commercial beacon devices (specifically, Estimote) at the construction site, with variations in field settings. The modifications included changes in 1) beacon installation height, 2) signal receiver position, and 3) the geometry of the indoor environment. Then, the signals were analyzed with box plots to quantitatively examine the impact of diverse conditions. Based on the results, an examination was conducted into the fundamental causes of beacon signal variations attributable to these factors. Concurrently, potential strategies to enhance the accuracy of beacon-based tracking and localization were proposed.

## 5.2 Literature Review On Construction Entitles Tracking and Localization Application

### 5.2.1 Tracking and localization technologies in construction

Previous research has proposed various methods for automatically obtaining the location data of construction entities. These can be broadly classified into 1) vision-based and 2) radio-based techniques. Vision-based approaches aim to locate and track construction entities through the analysis of 2D images from cameras or 3D data from laser scanning (Brilakis et al., 2010; Brilakis et al., 2011). For example, Yang et al. (2010) developed a 2D image-based tracking system to monitor the movements of construction workers. Also, Lee and Park (2019) employed a 3D stereo camera to track construction machinery and workers, further refining the tracking precision using the entity matching step. Though vision-based techniques offer precise tracking and localization with the added advantage of context, they require unobstructed views of the targets, making the methods susceptible to occlusions (Teizer, 2015). In active construction sites, the vulnerability becomes pronounced due to inevitable blind spots that arise from limited camera coverage, abundant site obstacles, and the consistent movement of entities (Zhang et al., 2020). So, employing the vision-based tracking and localizing construction objects presents challenges. Meanwhile, identity issues make the vision-based approach challenging in providing accurate tracking and locating information (Cai & Cai, 2020). Specifically, the IDs of tracking targets are difficult to associate correctly when the objects share similar shapes in the videos (e.g., multiple construction workers with the same uniforms and helmets) (Li et al., 2016), and they are also prone to switch when they are in close proximity (Zhang et al., 2021).

In contrast to the vision-based approach, the radio-based approaches exhibit fewer vulnerabilities to occlusions and ensure more dependable identity data. Such systems utilize signals transmitted between wireless communication devices to measure distance. Subsequently, the location is derived by processing distances through positioning algorithms such as the trilateration method (Brena et al., 2017). Various signal sources have been utilized in widely adopted radio-based tracking techniques, including Wi-Fi, Radio Frequency Identification (RFID), Ultra-Wideband (UWB), and Bluetooth (Fang et al., 2016). Among these, UWB is characterized by its short signal pulses and expansive bandwidth, which makes it less susceptible to site interferences such as walls, rebar meshes, and human interference. This ensures that UWB-based tracking maintains a commendable accuracy (Mahfouz & Kuhn, 2011). Notably, Maalek and Sadeghpour (2016) constructed a UWB Real-Time Location Estimation System (RTLS) in the lab and validated its accuracy to be less than 1.0 m for moving object tracking. UWB sensors have also been integrated with construction entities for real-time resource monitoring (Cheng & Teizer, 2013). However, such sensors demand significant infrastructure expenditure due to their limited signal range and high unit cost (Park, Kim, & Cho, 2016). The Wi-Fi-based positioning is another prominent alternative in construction. Woo et al. (2011) integrated Wi-Fi generators in a shield tunnel project. The process commenced with noise reduction using a filter algorithm, followed by the creation of a reference dataset. Location estimation was finalized by matching locations in this dataset. The researchers documented a positioning accuracy of five meters in an underground tunnel site in Guangzhou, China. However, Wi-Fi devices have continuous power demands, leading to frequent reallocations in ever-changing environments. Such adjustments result in a comprehensive calibration process (Obeidat et al., 2021), complicating their application in large-scale sites. Studies have also explored the potential of RFID-based tracking in construction. For instance, Cai et al. (2014) employed RFID to formulate a 3D location estimation algorithm for construction resources, achieving an accuracy of 2.5m. Additionally, Montaser and Moselhi (2014) proposed a cost-effective system for indoor location identification and material tracking. By installing RFID tags at reference positions, they ensured continuous monitoring of passing workers and materials, facilitating near-real-time tracking.

Of the myriad advancements in radio-based indoor positioning, Bluetooth low-energy (BLE) sensor stands out as the most apt for construction site applications due to the following advantages:

- Cost-efficiency: BLE sensors like Estimote provide a more affordable alternative than UWB and indoor GPS (Li & Becerik-Gerber, 2011). Their compatibility with commercial-grade smartphones, which adhere to the Bluetooth protocol, ensures a cost-effective tracking solution.
- Energy efficiency: BLE beacons, powered by button cells, have extended lifespans and demonstrate reduced energy consumption compared to Wi-Fi-based protocols.

- Flexible deployment: The compact design of BLE beacons, augmented by their button cell power sources, guarantees consistent performance and facilitates placement in ever-changing construction sites, obviating the need for a constant power supply (Fang et al., 2016).
- Data transmission convenience: Leveraging the iBeacon protocol, these sensors seamlessly connect to smartphones, enabling efficient message relays via Bluetooth.

Given these advantages, BLE beacons have found applications across various domains, as delineated in Table 5-1. Researchers have introduced beacon technology in construction, primarily for safety monitoring. For instance, Park, Kim and Cho (2016) showcased the practicality of a beacon-oriented positioning system by amalgamating it with the Building Information Model (BIM). Multiple strategies, such as fingerprinting and triangulation (Faragher & Harle, 2015; Martin et al., 2014), have been introduced to determine an object's location within the beacon signal mesh. Luo et al. (2011) also conducted tests on construction sites to validate the efficiency of different localization algorithms.

Table 5-1 Beacon applications review table

| Construction Stage | Method | Purpose | Research domain | Environment Setting | Related literature |
|---|---|---|---|---|---|
| Construction Phase | Localization | Tracking of construction entities (workers, materials, equipment) | Building information and construction | Outdoor | (Li et al., 2014; Lu et al., 2007; Luo et al., 2011; Park & Cho, 2017; Park et al., 2017; Teizer et al., 2020; Vähä et al., 2013; Zhao et al., 2019) |
| | | | | Indoor | (Park & Cho, 2017) |

| | | | | | |
|---|---|---|---|---|---|
| | | Safety alerts (construction equipment, fall detection, alarms) | Construction | Outdoor | (Baek & Choi, 2018; Gómez-de-Gabriel et al., 2019) |
| | Proximity Detection | Object tracking | Mobile sensing | Outdoor | (Ferreira et al., 2018) |
| | | Emergency safety assessment | Building information | Indoor | (Li et al., 2015) |
| | | Beacon signal evaluation | Electronics | Indoor | (Varsamou & Antonakopoulos, 2014) |
| Post-Construction Phase | Localization | Indoor tracking | Electronics | Indoor | (Chen et al., 2016; Dinh et al., 2020; Palumbo et al., 2015; Varsamou & Antonakopoulos, 2014) |
| | | Optimization of placement | Electronics | | (Rezazadeh et al., 2018) |
| | | Integration with Internet of Things systems | Electronics | | (He et al., 2015; Jeon et al., 2018; Ma & Cha, 2020) |
| | | Detection of indoor activities | Building information | | (Chen et al., 2019) |
| | Proximity detection | Indoor proximity-based tracking | Electronics | Indoor | (Zafari et al., 2017) |
| | | | Health | | (Kashimoto et al., 2017; Komai et al., 2016) |

| | | | | | |
|---|---|---|---|---|---|
| | | Detection of movement patterns | Mobile sensing | | (Vigneshwaran et al., 2015) |
| | | Augmentation of context-awareness in public transportation | Smart city (Transportation) | Outdoor | (Cianciulli et al., 2017) |

### 5.2.2 BLE beacon-based positioning methods

Among the Bluetooth-based positioning approaches (e.g., Angle of Arrival (AOA), Time of Arrival (TOA), Cell Identity (CI), Time Difference of Arrival (TDOA)), the Received (RX) power level-based mechanism stands out because of its flexibility of deployment and ease of calibration (Dimitrova et al., 2012). This method utilizes the signal propagation model. Within this framework, the received signal strength, quantified by the Received Signal Strength Indicator (RSSI) (Kotanen et al., 2003), exhibits an inverse correlation with the distance separating the beacon and the receiver (Kim et al., 2008). Such a relationship can be illustrated as follows in Equation (5-1)

$$\mathrm{RSSI} = -\,(10 \times n)\,\log 10(d) - a \tag{5-1}$$

In the equation, *n* represents a constant reflecting the propagation strength, *d* is the distance between the beacon and the receiver, and *a* stands for an offset value of RSSI (dBm) typically taken one meter from the beacon (Dong & Dargie, 2012). The RSSI values can be directly retrieved from the Bluetooth device's Host Controller Interface (HCI) (Bluetooth, 2001). Hence, once *n*, *d*, and *a* are determined, the estimated distance *d* can be computed by inserting the RSSI into the above formula (Pelant et al., 2017). However, determining the precise distance using this approach presents challenges. While Equation (5-1) depicts the Line-Of-Sight (LOS) propagation, which relates distance to the LOS path and the corresponding RSSI to the beacon's signal strength, it does not account for disruptions like multipath fading. The fading occurs when signals traverse multiple paths before reaching a receiver, which diminishes the accuracy of calculations (D. Chen et al., 2017). Additionally, the power of the received is considerably affected by environmental factors. Common materials like metal, water, concrete, and glass can degrade the RSSI value due to reflections and attenuations (Vieira et al., 2019). On construction sites, the continuous alterations of components, including rebar mesh and concrete walls, can adversely influence RSSI consistency, leading to imprecise and unreliable distance estimations.

Numerous studies have recommended the use of signal filters to mitigate signal degradation caused by multipath fading and environmental influences. For instance, Canton Paterna et al. (2017) applied Kalman filtering to reduce the noise signal and adopted the trilateration technique to pinpoint the position of stationary objects, achieving an accuracy of 0.7 meters. Wisanmongkol et al. (2019)  introduced a two-stage weighted average filtering technique that initially minimizes noise before amalgamating RSSI values from various BLE beacons. This method curtailed the estimation error to two meters. Furthermore, Mackey et al. (2020) employed three distinct Bayesian filtering methods to better align with the BLE signal distance equation, taking into account diverse error types. Their comparative tests showcased significant adjustments across two distinct settings, revealing a 30% boost in estimation accuracy. In a more recent development, Xu et al. (2021) introduced a real-time signal filtering approach, which yielded an average positional accuracy of approximately 0.8 meters. In addition to filtering signals, researchers have strived to simplify signal calibration using reference RSSI values. As noted by Caballero et al. (2008), the pre-calibrated relationship between received signal strength (RSSI) and distance is commonly adopted but often compromises the accuracy of distance estimations. Such models, often crafted by hardware manufacturers post-laboratory tests, typically follow the path loss formulation shown in Equation (5-1), with pre-determined parameters like the attenuation coefficient $n$ and offset $a$. Ma et al. (2017) extracted these from beacon manuals by matching them with specific receiver models, such as LG Nexus 5. However, the variability of RSSI due to environmental factors such as temperature, humidity, and signal interferences (e.g., walls, furnishings, human presence) questions the direct applicability of pre-calibrated models in target environments (Guidara et al., 2018). In this regard, recalibrating the RSSI-distance equation is essential before deploying beacons in a distinct environment. Dong and Dargie (2012) proposed two methods for this recalibration. The first method uses RSSI values, taken one meter from the beacons, as constant value *a* in the signal propagation model, subsequently determining the variable *n* with various RSSI-distance value pairs. The second method involves recording RSSI values by distance and performing curve-fitting approaches using the on-site data.

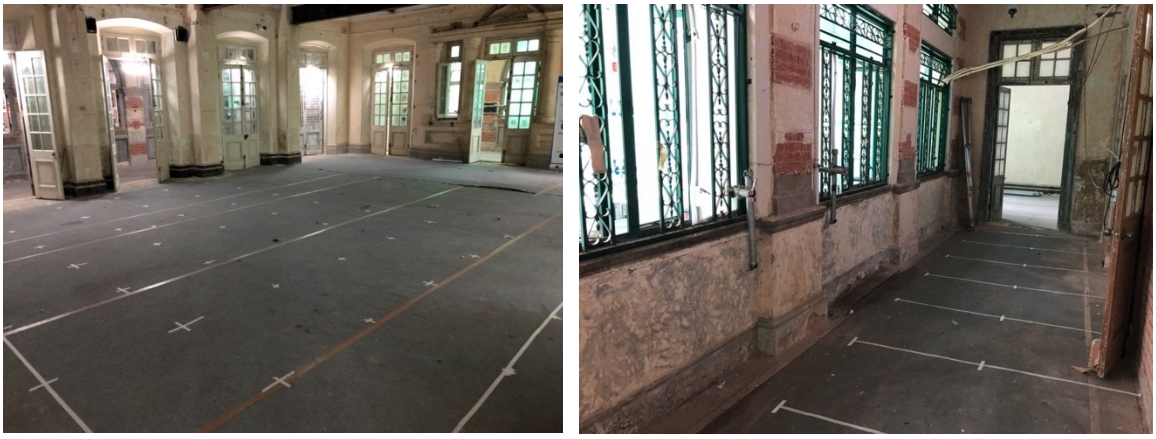
Establishing a reliable RSSI-distance relation is pivotal for determining the optimal worker localization model using BLE beacons, especially given the sensitivity of beacon signals. To achieve this goal, researchers assessed the reliability of pre-calibrated, calibrated, and curve-fitted models using actual data from construction sites. Concurrently, environmental factors potentially influencing the signal were scrutinized. Such analyses provided insights into the environmental impacts on the RSSI-distance relationship, directing refined beacon placements to reduce localization inaccuracies.

## 5.3 Field Data Collection

The research evaluated the reliability of pre-calibrated, calibrated, and curve-fitted models by utilizing RSSI data derived from BLE beacons on a construction site. Moreover, environmental variables such as the beacon's installation height, the signal receiver's condition, and the construction site's geometric layout were examined. This analysis illuminated the influence of environmental contexts on the RSSI in relation to distance. Such revelations informed more strategic beacon deployment aimed at minimizing signal vibration error. As a result, advancements in RSSI and beacon deployment methodologies were made, contributing to improved site localization reliability and stability.

In order to assess the signal performance of the BLE beacon in the construction site, the current study collected the beacon signal strength (i.e., RSSI) from the construction site located in St. Paul Secondary School, Hong Kong Island, Hong Kong SAR, managed by Able Engineering. The research spanned two indoor testbeds (Figure 5-1).Testbed 1 covers an area of 5.6 m × 8.5 m. The longer dimension, 8.5 m, corresponds to the distance between walls, while the 5.6 m side indicates an open space, underscoring the beacon deployment range. This arrangement includes eight beacons and 63 data collection points, as illustrated in Figure 5-2. Conversely, Testbed 2 is designed as a corridor, measuring 2.5 m in width and 8.5 m in length. The length represents the beacon deployment range and is bordered by open spaces at both ends. This setup accommodates seven beacons and includes 18 data acquisition points, as depicted in Figure 5-3. Beacons were installed across both testing arenas, with multiple units positioned within the designated testbeds. The research comprehensively covered the test areas, systematically recording signal intensities under various conditions. To log beacon signal strengths during the evaluations, an iOS software named "BLE Logger" was installed on an iPhone 7, serving as the primary instrument. This setup ensured each RSSI value was associated with a timestamp and hardware information, as detailed in Table 5-2.

During the data collection process, the smartphone used to capture signal strength readings was kept stationary at a predetermined location, 1.2 meters above the floor. Signal stability from the beacons, typically achieved within 30 seconds to one minute, prompted the iOS app to begin recording RSSI metrics from various beacons. Each trial concluded after the researcher navigated all designated data collection points, for example, the 63 points in Testbed 1, and logged the RSSI readings. Initially, beacons were placed at floor level (0.0 meters) before being elevated to 1.2 meters on columns to optimize signal reception. The dataset, collected from the experiment and comprising over 200,000 data points, originated from 12 distinct trials across two testbeds and involved 24 individual beacons.



(a) Testbed 1 Rectangular chamber (b) Testbed 2 Corridor

Figure 5-1 Site photos

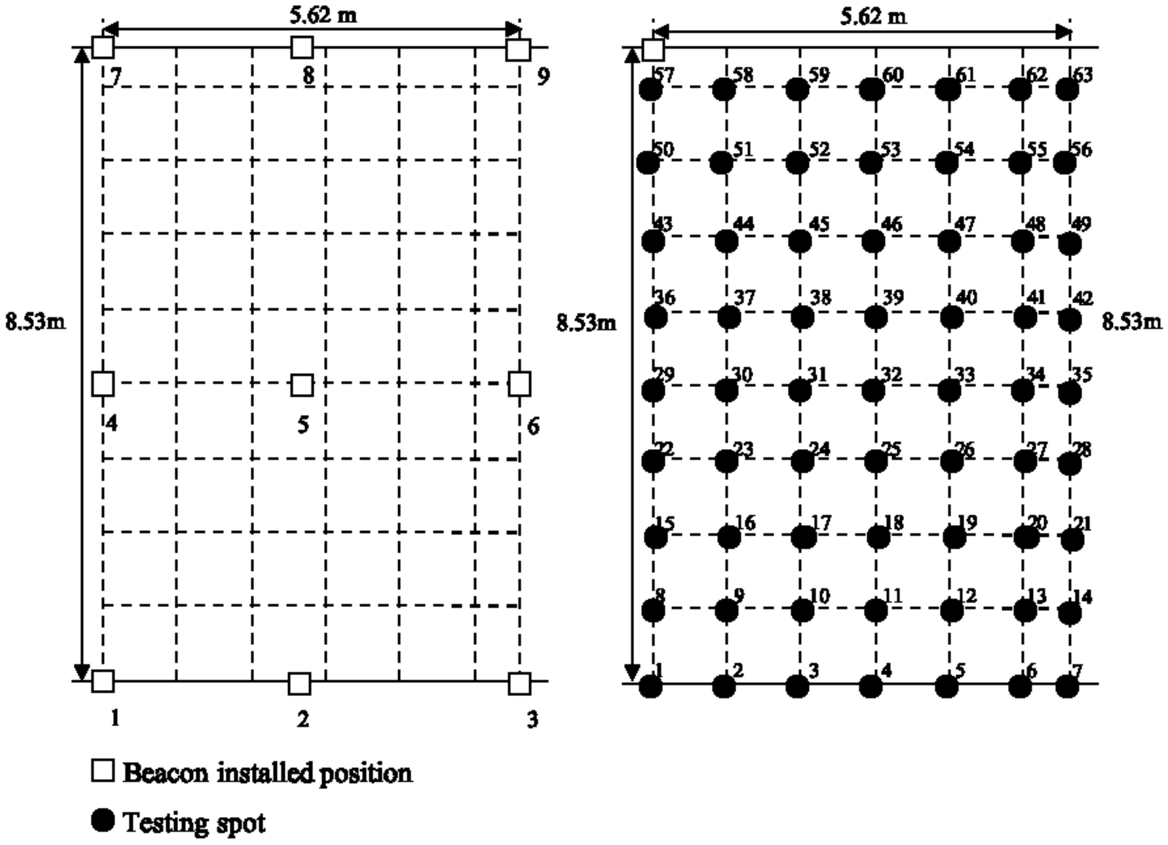


Figure 5-2 Testbed 1 Dimension and experiment setting

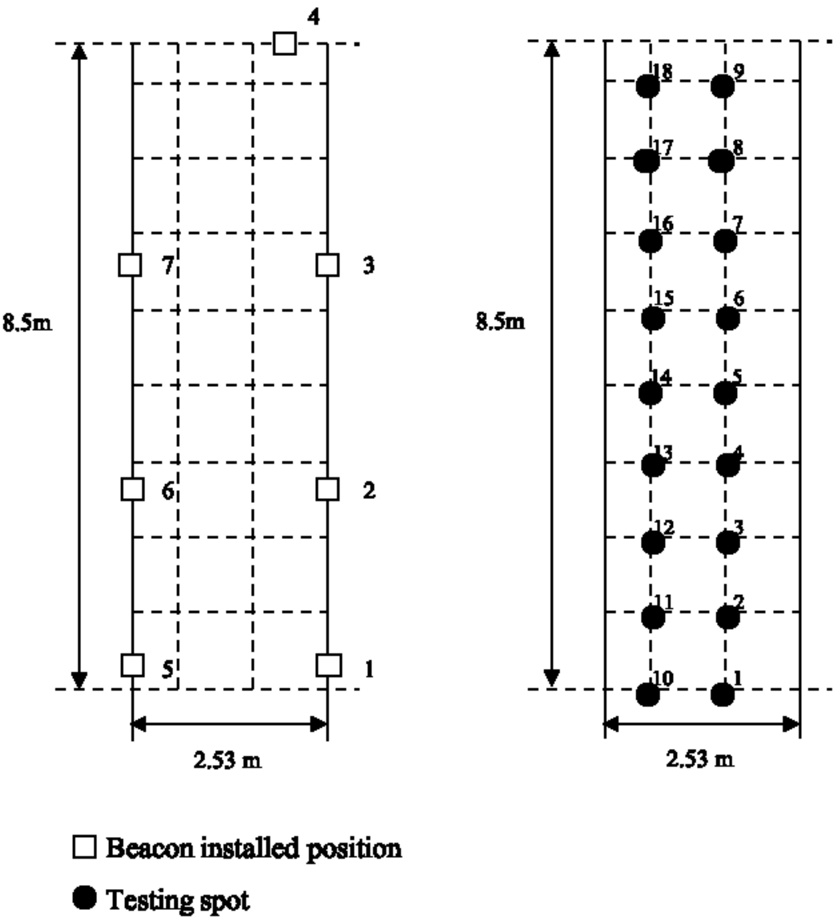


Figure 5-3 Testbed 2 Dimension and experiment setting

Table 5-2 Exhibit of raw data

| | |
|---|---|
| From: | example@connect.polyu.hk |
| Subject: | BLE Log History 722 |
| Date: | October 22, 2019, at 4:43:31 PM GMT+8 |
| Beacon UUID: | C504C2AD-F107-12BE-A7D3-E03DB541DB6F |

| Time | RSSI Value |
|---|---|
| 15/10/2019 18:11:30 | -82 |
| 15/10/2019 18:11:34 | -88 |
| 15/10/2019 18:11:37 | -87 |
| 15/10/2019 18:11:39 | -87 |
| 15/10/2019 18:11:40 | -87 |
| 15/10/2019 18:11:44 | -87 |

Based on the signal propagation theory (Seybold, 2005), the RSSI values expounded upon using the Log Distance Path Loss model, i.e., RSSI = $a \times \log(d) + b$. Parameters $a$ and $b$ encapsulate the compounded effects of both environmental and hardware factors (Goldhirsh & Vogel, 1998). Following the manufacturer's instructions, this study established the pre-calibrated line. Using the collected RSSI values and the distance between the beacon and the test spot, two more path loss models were identified: calibrated and fitted. The experiments also assessed the impact of height by comparing signal performance under two different scenarios. Furthermore, tests were conducted with the smartphone in two positions, inside and outside the pocket, to evaluate the influence of clothing texture.

Table 5-3 Beacons layout

| | Beacon spot | Beacon ID | |
|---|---|---|---|
| | | Wall (1.2 m)[1] | Floor (0 m) |
| Testbed 1 | 1 | Beacon 05 | Beacon08 |
| | 2 | Beacon 06 | Beacon09 |
| | 3 | Beacon 07 | Beacon10 |
| | 4 | Beacon 11 | |
| | 5 | Beacon 12 | |
| | 6 | Beacon 13 | Beacon21 |
| | 7 | Beacon 15 | Beacon14 |
| | 8 | Beacon 18 | Beacon16 |
| | 9 | Beacon 20 | Beacon19 |
| Testbed 2 | 1 | Beacon05 | Beacon18 |
| | 2 | Beacon07 | Beacon19 |
| | 3 | Beacon08 | Beacon20 |
| | 4 | Beacon10 | Beacon21 |
| | 5 | Beacon12 | Beacon22 |
| | 6 | Beacon13 | Beacon23 |
| | 7 | Beacon16 | Beacon24 |

Note[1]: Due to the sufficient number of beacons, two sets of beacons were installed at the same location, sharing identical coordinates on the ground. One set of beacons was installed on the wall (1.2 m high), while the other set was placed on the floor (0 m high).

Table 5-4 Experiment specifications

| Trial | Number of beacons | Beacon height | Receiver condition | Testbed |
|---|---|---|---|---|
| 1 | 8 | Floor (0 m) | Out of pocket | 1 |
| 3[1] | 8 | Column (1.2 m) | Out of pocket | 1 |
| 4 | 8 | Column (1.2 m) | In the pocket | 1 |
| 5 | 8 | Floor (0 m) | In the pocket | 1 |
| 7 | 16 | Column (1.2 m)<br>+ Floor (0 m)[2] | Out of pocket | 1 |
| 8 | 16 | Column (1.2 m)<br>+ Floor (0 m) | In the pocket | 1 |
| 9 | 7 | Column (1.2 m) | Out of pocket | 2 |
| 10 | 7 | Column (1.2 m) | In the pocket | 2 |
| 11 | 14 | Column (1.2 m)<br>+ Floor (0 m) [1] | Out of pocket | 2 |
| 12 | 14 | Column (1.2 m)<br>+ Floor (0 m) [1] | In the pocket | 2 |
| 13 | 7 | Floor (0 m) | Out of pocket | 2 |
| 14 | 7 | Floor (0 m) | In the pocket | 2 |

Note[1]: The data collected from Trials 2 and 6 were discarded due to their low quality. However, the data anticipated from Trials 2 and 6 were successfully collected during the follow-up data collection.

Note[2]: The researcher installed two sets of beacons (from the same manufacturer) in the same layout. One set of beacons was placed on the floor, while the other set was mounted on the wall (1.2 m high).

## 5.4 RSSI Data Analysis

The current study yielded three distinct models to describe the RSSI-distance relationship. Firstly, the pre-calibrated model represents the default relationship provided by the beacon manufacturer. Secondly, the calibrated model, derived from site-specific data, utilized RSSI values obtained at a 1.0-meter distance from the beacon and is articulated as RSSI = -12.6 log($d$) -68.9 dBm (Dong & Dargie, 2012). The third model, the fitted model, incorporates all the collected RSSI data from the site. This model is expressed as RSSI = -6.2 log($d$) -68.8 dBm through a comprehensive curve-fitting method, underscoring its derivation from an exhaustive dataset.

### 5.4.1 Signal strength and stability over distance

As shown in Figure 5-4, the pre-calibrated model exhibits a smaller RSSI (dBm) than the actual RSSI (dBm) from the site at given distances, particularly within 6 meters, indicating that the pre-calibrated model tends to underestimate distances at the provided RSSI (dBm) values. In contrast, the fitted model displays smaller RSSI (dBm) values at given distances compared to the calibrated model based on on-site data. Notably, the calibrated model, which uses actual RSSI (dBm) data from the site, presents the closest alignment with site data, especially within a range of 6 meters. These observations indicate potential errors when relying solely on the manufacturer's pre-calibrated model for estimating distances based on BLE beacons' RSSI (dBm) values. For reliable tracking and localization using BLE beacons, further calibration is advised.

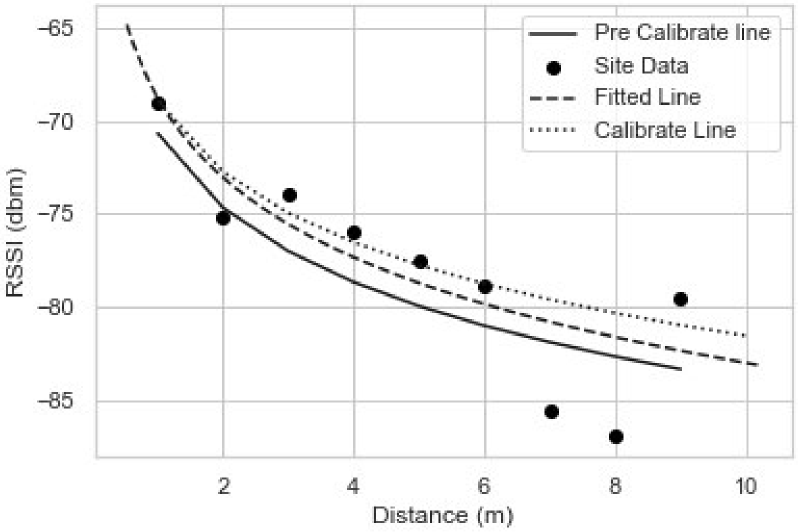
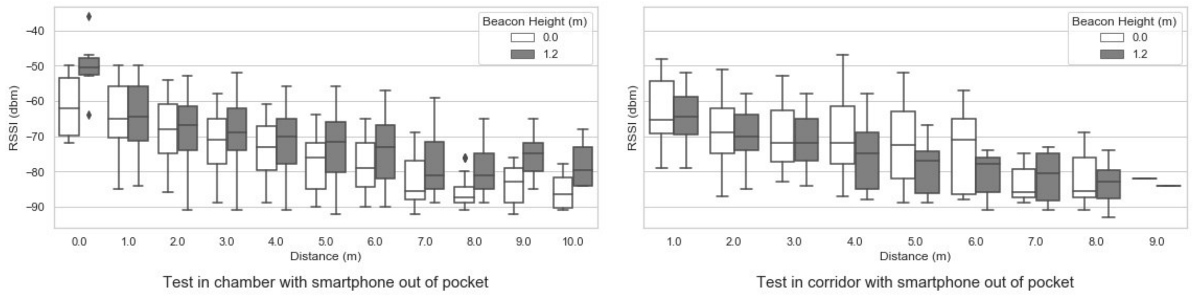


Figure 5-4 RSSI models for signal loss

### 5.4.2 Impact of beacon installation height

The comparative experiments were structured to discern the effect of installation height on the RSSI-distance relationship. Except for the variance in installation height, all other environmental variables, such as the beacon's location and signal acquisition point, were kept consistent. Notably, the receiving device was kept out of pocket to guarantee consistency in signal reception conditions. Then, two sets of RSSI signals were identified, each associated with varying beacon installation heights. The experiment investigated signal performance across two contrasting spatial environments, a rectangular room, and a corridor, to clarify how signals behave in different settings. The signal-distance relationship for the rectangular room is displayed in Figure 5-5 (a), and the data from the corridor is presented in Figure 5-5 (b). In both the rectangular room and the corridor, a t-test was conducted to assess the effect of installation height on RSSI values. The test compared RSSI readings from two heights in each setting. P-values were $5.5\times10^{-34}$ and 0.02 for the rectangular room and corridor, respectively. As both are below the 5.0% significance level, it indicates a significant influence of installation height on RSSI in both environments.

As presented in Figure 5-5, the signal performance conforms to the established signal propagation model across most of the plot. However, it is noted that at an elevation of 1.2 meters, RSSI values exhibit significant fluctuations when the receiver is close to the beacon, for example, within a distance of one meter. In contrast, beacons placed at floor level manifest minimal variance at short proximities. The RSSI readings become pronounced beyond an 8-meter expanse in a rectangular room. This finding is at odds with corridor data, where the placement of beacons at floor level is correlated with an RSSI amplification at a 5-meter threshold. However, such signal jump does not occur with beacons positioned at 1.2 meters. Accordingly, a specific effective range for RSSI can be determined where it follows the expected behavior of the signal propagation model. Upon analysis of the RSSI values within the prescribed effective distance, it is evident that signals captured at 1.2 meters and floor level exhibit a similar interquartile range aligned with the median in the rectangular space. Notably, the data from a height of 1.2 meters exhibit a broader range, suggesting a more significant variability and a higher likelihood of outliers. Conversely, in the corridor, the dataset obtained from the 1.2-meter elevation reveals a reduced interquartile range, indicative of enhanced signal stability and a diminished frequency of extreme data points.



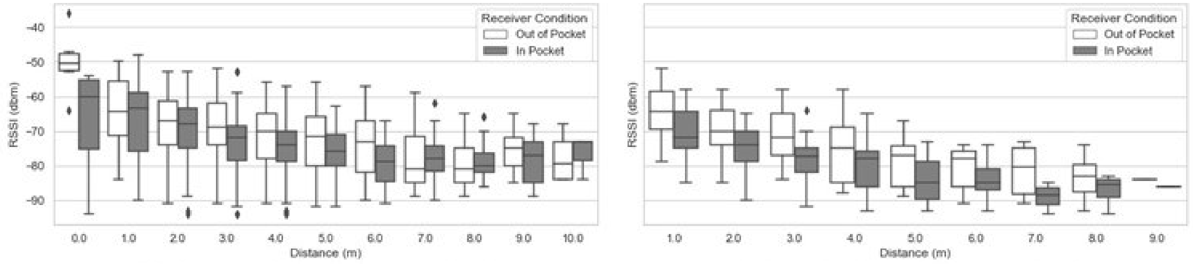
(a) Signal performance in the rectangular room (b) Signal performance in the corridor

Figure 5-5 RSSI- distance plot with the different beacon installation height

### 5.4.3 Impact of receiver condition

Consistent variables such as beacon installation height and the geometry of the indoor environment were controlled to investigate the impact of user interference on RSSI values. Specifically, the influence of smartphone positioning, whether kept in-pocket or out-of-pocket, was analyzed. It is critical to note that the smartphone's height was maintained at 1.2 meters during the experiment. The RSSI data were represented as box plots in Figure 5-6, with the left graph illustrating readings from a rectangular room (testbed 1) and the right graph depicting data from a corridor (testbed 2). Expecting the RSSI values to exhibit a normal distribution due to the extensive dataset, a t-test was conducted to evaluate the significance of the variation caused by the smartphone being in a pocket. The p-values obtained from the RSSI data for the rectangular room and the corridor were $6.2\times10^{-11}$ and $1.9\times10^{-18}$, respectively. Given that both values fall well below the 5.0% significance level, the data suggests that the smartphone in a pocket exerts a considerable effect on RSSI readings within the specific indoor environments of a construction site.

As depicted in Figure 5-6, the signal adheres to the expected signal propagation law within a specific effective range, analogous to the patterns depicted in Figure 5-5. The stable range for RSSI signals extends from one to seven meters in the rectangular room, whereas in the corridor, the consistent RSSI-distance relationship is observed from one meter up to six meters. The signal strength demonstrates significant variability when assessed outside this established range, including at nearer and farther proximities. The dataset analysis, notably when the smartphone was enclosed in a pocket, showed an increased frequency of outliers. However, the central data points were characterized by a narrower interquartile range, illustrating a higher level of consistency in signal reception among the core data points.



Test in chamber with installing Beacon at 1.2 m high

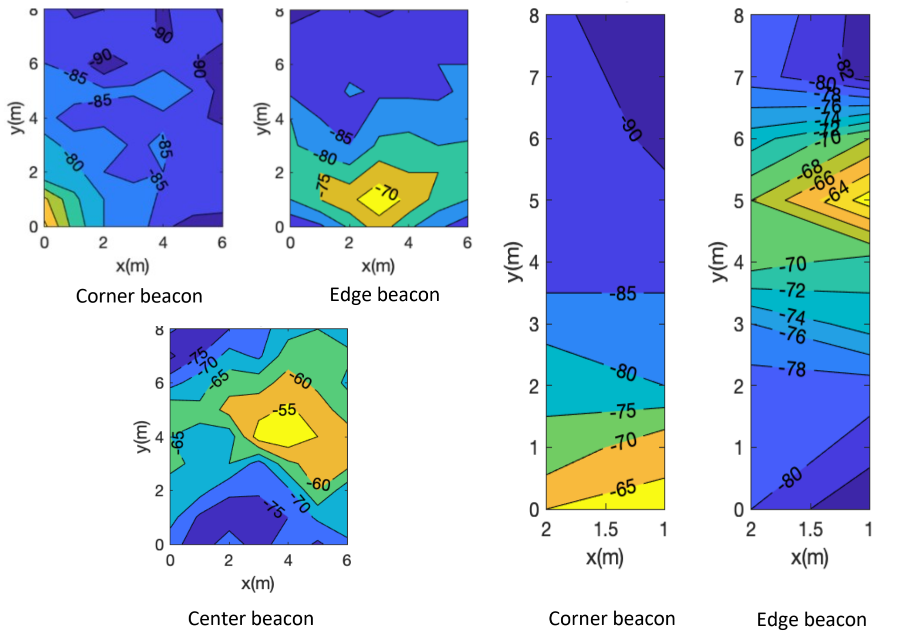
Test in corridor with installing Beacon at 1.2 m high

Figure 5-6 RSSI- distance plot under different receiver conditions

### 5.4.4 Impact of environment geometry

Figure 5-7 presents contour plots of beacon signal strength with the signals collected by beacons installed at a height of 1.2 meters. Subplot (a) depicts results from the rectangular room, while subplot (b) portrays signal distributions within the corridor. Within these plots, points on the same contour curve possess equivalent signal strength values, as denoted on the curves. These contour intervals remain consistent, with color variations representing signal strength, transitioning from light to dark as the RSSI diminishes, indicating weaker signal strength. Based on their placement, beacons were divided into three categories: center, edge, and corner beacons. Center beacons were strategically positioned at the geometric center of the floor, distant from the walls. In contrast, edge beacons were placed proximal to a single wall, and corner beacons were situated near the doorway corners.

Oval-like or serrated-like contour lines within the signal contour plot signify a uniform signal distribution, typically observed in proximity to the beacon. However, as distance increases, these contours begin to deviate into more irregular shapes. As illustrated in Figure 5-7, the center beacon in the rectangular room exhibits the most consistent and uniform signal contour compared to its edge and corner counterparts. The uniform distribution of the rectangular room's corner beacon concludes when the RSSI value drops below -80 dBm, mirroring the edge beacon's pattern. Nevertheless, the coverage of the stable signal from the corner beacon is approximately half that of the edge beacon. Only corner and edge beacons were installed within the corridor's narrower confines. Signal distributions between these beacon types appeared relatively comparable, though the contour intervals for the edge beacon were noticeably tighter than those of the corner beacon.



(a) Signal contours in a rectangular room

(b) Signal contours in a corridor

Figure 5-7 Beacon signal contours map

## 5.5 Discussion

Utilizing a pre-calibrated model for distance estimation is advised in practice as it mitigates the extensive workload associated with calibration (Barsocchi et al., 2009). Nevertheless, the distance errors using the pre-calibrated line are more significant than those observed with the fitted and calibrated lines (Figure 5-4). While indicative of the signal-distance relationship specific to the actual site, the fitted model necessitates considerable data acquisition efforts. Given these demands, the curve-fitting approach is not recommended for beacon signal calibration in construction sites. In contrast, the calibrated model, though necessitating less setup effort, demonstrates a robust congruence with the site-specific data, surpassing the performance of the fitted model. The calibrated model is endorsed for subsequent tracking and localization endeavors.

The subsequent analysis indicates that signals do not consistently adhere to the propagation model, as depicted in Figure 5-6 and Figure 5-6. The effective distance range, wherein the RSSI measurements comply with the signal propagation law, spans from one to seven meters in the rectangular room and one to six meters in the corridor. Beyond these effective ranges, significant variations in beacon signal strength occur when the beacon is either exceedingly close to or far from the receiver. Accordingly, to accurately utilize the model for distance estimation, the beacons should be positioned within the ascertained effective range relative to the target, tailored to the specific environmental context.

Comparative analysis of RSSI performance at varying installation heights reveals that the placement of beacons significantly affects the stability of RSSI readings and, consequently, the accuracy of distance estimation. The statistical outcomes highlight the significant effect of ground reflections on RSSI stability. As demonstrated in Figure 5-5, deploying the beacon at floor level in more expansive spaces results in compact data distribution, evidenced by the condensed box plot. Such configuration implies diminished variability and improved signal stability, crucial for reliable distance estimation amidst signal fluctuations. Conversely, in more confined spaces such as corridors, the result supports a preference for beacon placement at elevated heights, where RSSI measurements display a smaller spread, enhancing measurement consistency and reliability.

Furthermore, placing a smartphone in a pocket statistically significantly affects the RSSI. Comparative analysis reveals that the incidence of outlier measurements increases when the signal receiver is enclosed within a pocket. Despite this, the central tendency of the data remains robust, with readings aggregating more densely around the median for both environments under study. In applications such as tracking construction workers, the convenience of pocket storage for smartphones is unavoidable. In these instances, using sophisticated signal filters becomes crucial to mitigate the effects of outliers, thereby ensuring a more uniform and dependable dataset. However, filtering signals outlier incurs a response delay. In situations requiring swift reaction and minimal outlier frequency, particularly within hazardous zone alarm systems, it is advised not to place receivers in pockets to avoid delays and preserve data credibility.

The current study then mapped signal contours to analyze signal distribution across varied environments. Figure 5-7 illustrates that beacons positioned at corners and edges exhibit more uniform and consistent contour shapes, denoting stable RSSI-distance relationships within a short range, particularly within a proximate radius of up to four meters. Beacons at corners gain the advantage of two walls serving as reflective surfaces, while those at edges benefit from a single wall. In contrast, a centrally located beacon relies solely on the ground as a reflective surface. The findings elucidate that proximity to multiple reflective planes correlates with more uniform contour patterns close to the beacon, signifying a dependable signal strength gradient with incremental distances, thereby enhancing distance estimation accuracy. For open areas such as Testbed 1, installing beacons at corners or edges is recommended to ensure a reliable RSSI distance relationship within a four-meter range. Central beacons, however, are not advised due to their less dependable performance. In the absence of walls, establishing manual solid reflective planes near the beacon may stabilize the signal, ensuring that reflections near the beacon predominate. In environments lacking such controlled conditions, multiple factors influence the signal, leading to a complex interplay that complicates the characterization of signal behavior. Conversely, in confined spaces like the corridor in Testbed 2, the signal contour exhibits a stable and homogenized signal gradient, offering the potential for distance estimation using RSSI. The advised range for employing this model is limited to four meters.

The study's recommendation for the calibrated model, while minimizing setup efforts, is not without constraints, as it requires a careful balance between practicality and the precision of site-specific calibrations. Additionally, the delineated effective distance range for beacon placement implies a limitation in spatial application, particularly in areas beyond the identified range. To mitigate this limitation, deploying multiple beacons at intervals within the effective range determined by the study is recommended. The investigation also brings to light the limitations of RSSI stability with varied environmental factors, such as the impact of smartphone storage on signal quality, which could affect the practical deployment of RSSI-based tracking systems. However, the current research evaluates three significant factors influencing signal stability and does not encompass a comprehensive list of potential variables. Future studies should aim to identify additional influencing factors. Furthermore, the current experiment does not incorporate mobile objects or construction entities within the test environment, which could affect signal behavior. Consequently, ensuing research could enhance the validity of findings by including such dynamic elements in the testbed configuration, thereby providing a more robust assessment of RSSI stability across varied environmental conditions and contributing to the development of more resilient RSSI-based tracking systems.

## 5.6 Conclusion

The current study aims to validate the feasibility of using BLE beacon-based systems for tracking and locating within construction sites. Moreover, it seeks to delineate the effects of diverse factors on the RSSI-distance model, thereby formulating deployment guidance for practical application in construction environments. This study examined three path loss models: the pre-calibrated model, the calibrated model, and the fitted model. The latter two, derived from the gathered RSSI values and the corresponding distances between the beacon and the test locations, provided a more customized signal propagation assessment. According to the results, the calibrated model is recommended for tracking and localization in construction sites because it combines robust performance with reduced setup effort, offering a practical and accurate solution. This model is described by the equations RSSI = -12.6 log(*d*) -68.9 dBm. The comprehensive test results also reveal that the RSSI maintains a stable signal path loss model exclusively within a specific distance range. Such stability is observed from one to four meters in the construction environment.

Utilizing a calibrated model for distance estimation necessitates environment-specific adaptations when implementing such models. Consequently, the present study validates the environmental parameters affecting signal distance models' stability and precision. The statistical testing reveals that the height at which beacons are installed above the ground and the texture of the surrounding environment significantly influence signal reception. This study recommends placement strategies tailored to different environments to establish a BLE beacon network capable of generating a stable signal mesh for accurate distance estimation. In open areas, positioning beacons at ground level is preferable. Conversely, elevating the beacon to a specific height in confined spaces proves to be more effective. When the receiver must be carried in a pocket, the data indicate an increased likelihood of signal anomalies, which are unsatisfactory for applications requiring immediate response, such as alarm systems. Moreover, installing beacons adjacent to sturdy and smooth surfaces, like walls, can facilitate a stable RSSI-distance relationship within a short range, typically between one to four meters, where reflections are predominant.

In conclusion, the current research delineates the feasibility and challenges of implementing beacon signal-based tracking and localization systems within indoor construction environments. The findings are crucial for optimizing beacon placement, thereby establishing a more stable signal network. This groundwork paves the way for future research endeavors to explore various influencing factors and more complex scenarios.

# CHAPTER 6 CONCLUSIONS AND FUTURE RESEARCH

The construction industry, characterized by its intensive reliance on manual labor, faces significant productivity challenges due to this dependence. Task-level activity analysis is crucial for identifying the root causes of low productivity issues, involving monitoring and categorization of activities based on their impact on productivity. The integration of advanced sensing technologies has significantly improved the process of monitoring construction activities by facilitating the collection of activity and location data in real time. Leveraging machine learning and deep learning algorithms, researchers can construct models for recognizing actions and tracking locations, thereby analyzing task-level activity in an automated, accurate, and timely manner and also offering a comprehensive, data-driven insight into operational efficiencies. Despite the promising potential of sensor-based construction task-level activity analysis, challenges related to data interpretability, concerns about the accuracy and reliability of sensor-based methods, and the complexities of implementing these technologies in unstructured environments like construction sites are significant barriers to their widespread application.

Therefore, the current study developed four steps to address the challenges and complete the application framework. As shown in Figure 1-1, a hierarchical work taxonomy tailored for task-level activity analysis was designed, laying the groundwork for the objective evaluation of construction tasks. This taxonomy, essential for the analysis, emphasized the productivity potential of activities and the distinctiveness of body movements, enhancing data interpretability and reducing the likelihood of misclassification in action recognition.

Subsequently, the practical applicability of the work taxonomy was validated through rigorous field experiments involving eighteen construction workers across two sites in Hong Kong, focusing on concrete work tasks and formwork tasks, which are essential and representative works in construction. Utilizing an inertial measurement unit (IMU) in a smartwatch, acceleration data were collected and analyzed, applying both traditional and advanced machine learning algorithms for action recognition. This validation not only tested the taxonomy's feasibility but also its effectiveness in real-world conditions. The research findings indicate that the proposed taxonomy effectively classifies construction activities with high accuracy for Level 1 (above 95%) and acceptable accuracy for Level 2 (74.6% to 83.8%), demonstrating the taxonomy's capability to convey comprehensive activity information and handle noisy data. Additionally, the study successfully estimated the duration of activities in Level 2, allowing for the assessment of work efficiency and identification of causes behind low productivity, thereby facilitating potential improvements in construction tasks.

Further advancing the study, a novel sensor fusion approach for action recognition was introduced, integrating image and acceleration data to leverage the unique strengths of each modality. Through laboratory experiments, the complementary benefits of decision-level fusion approaches in recognizing construction activities were demonstrated, underscoring the enhanced accuracy and reliability achieved by combining data from diverse sensors. The study developed four decision-level fusion methods, specifically the Dempster-Shafer (DS), Weighted Dempster-Shafer (WDS), Topk Weighted Dempster-Shafer (TopkWDS), and Thresholding Weighted Dempster-Shafer (TWDS), significantly outperformed individual acceleration and video-based models in construction action recognition, with TWDS achieving the highest accuracy of 85.67%, marking a 13.9% and 4.43% increase compared to solely using video and acceleration models, respectively.

Lastly, the application of BLE beacon-based localization under various site conditions was explored to better understand the principles of deploying BLE beacons for construction site localization. The investigation into different environmental setups and beacon configurations provided valuable insights into optimizing sensor deployment for accurate and efficient site monitoring. The study confirms the effectiveness of BLE beacon-based systems for tracking and localization in construction sites, particularly advocating for the calibrated model due to its balance of robust performance and ease of setup. The model derived from the site's collected data is RSSI = -12.6 log(*d*) -68.9 dBm. It highlights the impact of environmental factors like beacon height and surrounding textures on signal accuracy, recommending environment-specific beacon placement for stable signal transmission.

The study also highlights limitations in action recognition for construction activities using acceleration signals, especially at Level 3, where classification accuracy significantly decreases due to the complexity of micro-level analysis and the challenge of distinguishing between similar movements. Future research should aim to enhance the transition between activities, optimize time-series data segmentation for more accurate classification, and employ additional techniques, such as dynamic and fuzzy segmentation, to manage continuous and overlapping activity data better. Furthermore, the study suggests further validation of the proposed work taxonomy and action recognition algorithms with more extensive field data to improve the reliability and practical applicability in construction settings. It also acknowledges the challenges in accurately predicting specific activities like "Idling" despite improvements from decision-level fusion and filtered weights. Future efforts will focus on improving data collection in both laboratory and real-world environments, involving more participants with diverse backgrounds to enhance model generalization, and employing Leave-One-Subject-Out Validation (LOSOV) to address individual variances, making the developed framework more applicable to various construction scenarios. Additionally, future research will explore the resilience and flexibility of decision-level fusion under different conditions and with various sensor types to ensure robustness against sensor failures and environmental variations. The study on BLE beacon signal in-site validation endorses the calibrated model as the preferred approach for beacon signal calibration in construction sites due to its minimal setup requirements and strong performance despite data limitations and calibration challenges. It also notes that the effective distance range for beacon placement is limited, impacting RSSI stability and accuracy in distance estimation, with factors like smartphone storage affecting signal quality. Future research will further investigate factors influencing RSSI stability, including the impact of mobile objects in the environment, to develop more dependable RSSI-based tracking systems.

In conclusion, this study contributes to the field of construction management by advancing the understanding and application of sensor-based activity analysis and localization techniques. Through rigorous methodology and comprehensive data analysis, this work enhances the exposure of causes for low productivity issues in construction projects. Furthermore, the integration of advanced technologies and data-driven approaches will continue to play a pivotal role in overcoming industry challenges and achieving operational excellence.

# REFERENCE

Abidi, M. A., & Gonzalez, R. C. (1992). *Data fusion in robotics and machine intelligence*. Academic Press Professional, Inc.

Abonyi, J., Feil, B., Nemeth, S., & Arva, P. (2005). Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series. *Fuzzy Sets and Systems*, *149*(1), 39-56.

Adrian, J. J., & Boyer, L. T. (1976). Modeling method-productivity. *Journal of the Construction Division*, *102*(1), 157-168.

Akhavian, R., & Behzadan, A. H. (2016). Smartphone-based construction workers' activity recognition and classification. *Automation in Construction*, *71*, 198-209.

Amundson, I., Kusy, B., Volgyesi, P., Koutsoukos, X., & Ledeczi, A. (2008). Time synchronization in heterogeneous sensor networks. International Conference on Distributed Computing in Sensor Systems,

Ashry, S., Ogawa, T., & Gomaa, W. (2020). CHARM-deep: Continuous human activity recognition model based on deep neural network using IMU sensors of smartwatch. *IEEE Sensors Journal*, *20*(15), 8757-8770.

Ayed, S. B., Trichili, H., & Alimi, A. M. (2015). Data fusion architectures: A survey and comparison. 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA),

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481-2495.

Badrinath, G., & Gupta, P. (2009). Feature level fused ear biometric system. 2009 Seventh International Conference on Advances in Pattern Recognition, Baek, J., & Choi, Y. (2018). Bluetooth-Beacon-Based Underground Proximity Warning System for Preventing Collisions inside Tunnels. *Applied Sciences*, *8*(11), 2271. https://doi.org/10.3390/app8112271

Bahrepour, M., Meratnia, N., Taghikhaki, Z., & Havinga, P. (2011). Sensor fusion-based activity recognition for Parkinson patients. *Sensor Fusion-Foundation and Applications*, 171-190.

Bangaru, S. S., Wang, C., Busam, S. A., & Aghazadeh, F. (2021a). ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors. *Automation in Construction*, *126*, 103653.

Bangaru, S. S., Wang, C., Busam, S. A., & Aghazadeh, F. J. A. i. C. (2021b). ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors. *126*, 103653.

Banos, O., Galvez, J.-M., Damas, M., Pomares, H., & Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, *14*(4), 6474-6499.

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. International conference on pervasive computing,

Barsocchi, P., Lenzi, S., Chessa, S., & Giunta, G. (2009). Virtual calibration for RSSI-based indoor localization with IEEE 802.15. 4. 2009 IEEE International Conference on Communications,

Benali Amjoud, A., & Amrouch, M. (2020). Convolutional neural networks backbones for object detection. International Conference on Image and Signal Processing,

Benediktsson, J. A., & Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(3), 1367-1377.

Beni, G., Hackwood, S., Hornak, L., & Jackel, J. (1983). Dynamic sensing for robots: an analysis and implementation. *The International Journal of Robotics Research*, *2*(2), 51-61.

Blanke, U., & Schiele, B. (2010). Remember and transfer what you have learned-recognizing composite activities based on activity spotting. International Symposium on Wearable Computers (ISWC) 2010,

Blasch, E., Valin, P., & Bosse, E. (2010). Measures of effectiveness for high-level fusion. 2010 13th International Conference on Information Fusion,

Blasch, E. P., & Plano, S. (2002). JDL Level 5 fusion model: user refinement issues and applications in group tracking. Signal processing, sensor fusion, and target recognition XI,

Bohn, J. S., & Teizer, J. (2010). Benefits and barriers of construction project monitoring using high-resolution automated cameras. *Journal of construction engineering and management*, *136*(6), 632-640.

Bowman, C., & Morefield, C. (1980). Multisensor fusion of target attributes and kinematics. 1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes,

Brena, R. F., García-Vázquez, J. P., Galván-Tejada, C. E., Muñoz-Rodriguez, D., Vargas-Rosales, C., & Fangmeyer, J. (2017). Evolution of indoor positioning technologies: A survey. *Journal of Sensors*, *2017*.

Brilakis, I., Lourakis, M., Sacks, R., Savarese, S., Christodoulou, S., Teizer, J., & Makhmalbaf, A. (2010). Toward automated generation of parametric BIMs based on hybrid video and laser scanning data. *Advanced Engineering Informatics*, *24*(4), 456-465.

Brilakis, I., Park, M.-W., & Jog, G. (2011). Automated vision tracking of project related entities. *Advanced Engineering Informatics*, *25*(4), 713-724.

Buchholz, B., Paquet, V., Wellman, H., & Forde, M. (2003). Quantification of ergonomic hazards for ironworkers performing concrete reinforcement tasks during heavy highway construction. *AIHA journal*, *64*(2), 243-250.

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, *46*(3), 1-33.

Caballero, F., Merino, L., Maza, I., & Ollero, A. (2008). A particle filtering method for wireless sensor network localization with an aerial robot beacon. 2008 IEEE International Conference on Robotics and Automation,

Cai, H., Andoh, A. R., Su, X., & Li, S. (2014). A boundary condition based algorithm for locating construction site objects using RFID and GPS. *Advanced Engineering Informatics*, *28*(4), 455-468.

Cai, J., & Cai, H. (2020). Robust Hybrid Approach of Vision-Based Tracking and Radio-Based Identification and Localization for 3D Tracking of Multiple Construction Workers. *Journal of Computing in Civil Engineering*, *34*(4). https://doi.org/10.1061/(asce)cp.1943-5487.0000901

Canton Paterna, V., Calveras Auge, A., Paradells Aspas, J., & Perez Bullones, M. A. (2017). A Bluetooth Low Energy Indoor Positioning System with Channel Diversity, Weighted Trilateration and Kalman Filtering. *Sensors (Basel)*, *17*(12), 2927. https://doi.org/10.3390/s17122927

Castanedo, F. (2013). A review of data fusion techniques. *The scientific world journal*, *2013*.

Chapman, R. E., Butry, D. T., & Huang, A. L. (2010). Measuring and improving US construction productivity. Proceedings of TG65 and W065-Special Track. 18th CIB World Building Congress, Chen, C., Jafari, R., & Kehtarnavaz, N. (2014). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Transactions on Human-Machine Systems*, *45*(1), 51-61.

Chen, C., Jafari, R., & Kehtarnavaz, N. (2017). A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, *76*(3), 4405-4425.

Chen, D., Shin, K. G., Jiang, Y., & Kim, K.-H. (2017). Locating and tracking ble beacons with smartphones. Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies,

Chen, H., Cha, S. H., & Kim, T. W. (2019). A framework for group activity detection and recognition using smartphone sensors and beacons. *Building and Environment*, *158*, 205-216. https://doi.org/10.1016/j.buildenv.2019.05.016

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, *54*(4), 1-40.

Chen, Z., Zhu, Q., & Soh, Y. C. (2016). Smartphone Inertial Sensor-Based Indoor Localization and Tracking With iBeacon Corrections. *IEEE Transactions on Industrial Informatics*, *12*(4), 1540-1549. https://doi.org/10.1109/tii.2016.2579265

Cheng, T., & Teizer, J. (2013). Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. *Automation in Construction*, *34*, 3-15.

Cheng, T., Teizer, J., Migliaccio, G. C., & Gatti, U. C. (2013). Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data. *Automation in Construction*, *29*, 24-39.

Cianciulli, D., Canfora, G., & Zimeo, E. (2017). Beacon-based context-aware architecture for crowd sensing public transportation scheduling and user habits. *Procedia Computer Science*, *109*, 1110-1115. https://doi.org/10.1016/j.procs.2017.05.451

Cobb, B. R., & Shenoy, P. P. (2003). A comparison of Bayesian and belief function reasoning. *Information Systems Frontiers*, *5*(4), 345-358.

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, *29*.

Dasarathy, B. V. (1994). *Decision fusion* (Vol. 1994). IEEE Computer Society Press Los Alamitos.

Datar, M., Gionis, A., Indyk, P., & Motwani, R. (2002). Maintaining stream statistics over sliding windows. *SIAM journal on computing*, *31*(6), 1794-1813.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning,

De Boer, R. C. (2002). A Generic architecture for fusion-based intrusion detection systems Citeseer].

Dempster, A. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics*, *38*(2), 325-339.

Dietterich, T. G. (2002). Machine learning for sequential data: A review. Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR),

Dimitrova, D. C., Alyafawi, I., & Braun, T. (2012). Experimental comparison of bluetooth and wifi signal propagation for indoor localisation. International Conference on Wired/Wireless Internet Communications, Ding, Y., Yao, X., Wang, S., & Zhao, X. (2019). Structural damage assessment using improved Dempster-Shafer data fusion algorithm. *Earthquake Engineering and Engineering Vibration*, *18*, 395-408.

Dinh, T.-M. T., Duong, N.-S., & Sandrasegaran, K. (2020). Smartphone-Based Indoor Positioning Using BLE iBeacon and Reliable Lightweight Fingerprint Map. *IEEE Sensors Journal*, *20*(17), 10283-10294. https://doi.org/10.1109/jsen.2020.2989411

Dong, Q., & Dargie, W. (2012). Evaluation of the reliability of RSSI for indoor localization. 2012 International Conference on Wireless Communications in Underground and Confined Areas,

Dubois, D., & Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational intelligence*, *4*(3), 244-264.

Dvornik, N., Shmelkov, K., Mairal, J., & Schmid, C. (2017). Blitznet: A real-time deep network for scene understanding. Proceedings of the IEEE international conference on computer vision,

Dzeng, R.-J., Lin, C.-W., & Hsiao, F.-Y. (2014). Application of RFID tracking to the optimization of function-space assignment in buildings. *Automation in Construction*, *40*, 68-83.

Elmenreich, W. (2002). An introduction to sensor fusion. *Vienna University of Technology, Austria*, *502*, 1-28.

Elnahrawy, E., Li, X., & Martin, R. P. (2004). The limits of localization using signal strength: A comparative study. 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.,

Esteban, J., Starr, A., Willetts, R., Hannah, P., & Bryanston-Cross, P. (2005). A review of data fusion models and architectures: towards engineering guidelines. *Neural Computing & Applications*, *14*(4), 273-281.

Everett, J. G., & Slocum, A. H. (1994). Automation and robotics opportunities: construction versus manufacturing. *Journal of Construction Engineering and Management*, *120*(2), 443-452.

Fang, Y., Cho, Y. K., Zhang, S., & Perez, E. (2016). Case Study of BIM and Cloud–Enabled Real-Time RFID Indoor Localization for Construction Management Applications. *Journal of construction engineering and management*, *142*(7), 05016003. https://doi.org/10.1061/(asce)co.1943-7862.0001125

Faragher, R., & Harle, R. (2014). An analysis of the accuracy of bluetooth low energy for indoor positioning applications. Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014),

Faragher, R., & Harle, R. (2015). Location Fingerprinting With Bluetooth Low Energy Beacons. *IEEE journal on Selected Areas in Communications*, *33*(11), 2418-2428. https://doi.org/10.1109/jsac.2015.2430281

Ferreira, J. C., Resende, R., & Martinho, S. (2018). Beacons and BIM Models for Indoor Guidance and Location. *Sensors (Basel)*, *18*(12), 4374. https://doi.org/10.3390/s18124374

Forde, M. S., & Buchholz, B. (2004). Task content and physical ergonomic risk factors in construction ironwork. *International Journal of Industrial Ergonomics*, *34*(4), 319-333.

Goldhirsh, J., & Vogel, W. J. (1998). Handbook of propagation effects for vehicular and personal mobile satellite systems. *NASA Reference Publication*, *1274*, 40-67.

Golestani, N., & Moghaddam, M. (2020). Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. In: Google Patents.

Gómez-de-Gabriel, J. M., Fernández-Madrigal, J. A., López-Arquillos, A., & Rubio-Romero, J. C. (2019). Monitoring harness use in construction with BLE beacons. *Measurement*, *131*, 329-340.

Gong, Y., Yang, K., Seo, J., & Lee, J. G. (2022). Wearable acceleration-based action recognition for long-term and continuous activity analysis in construction site. *Journal of Building Engineering*, *52*, 104448.

Gouett, M. C., Haas, C. T., Goodrum, P. M., & Caldas, C. H. (2011). Activity analysis for direct-work rate improvement in construction. *Journal of Construction Engineering and Management*, *137*(12), 1117-1124.

Gu, T., Wu, Z., Tao, X., Pung, H. K., & Lu, J. (2009). epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. 2009 IEEE International Conference on Pervasive Computing and Communications,

Guidara, A., Fersi, G., Derbel, F., & Jemaa, M. B. (2018). Impacts of temperature and humidity variations on RSSI in indoor wireless sensor networks. *Procedia Computer Science*, *126*, 1072-1081.

Gumaei, A., Hassan, M. M., Alelaiwi, A., & Alsalman, H. J. I. A. (2019). A hybrid deep learning model for human activity recognition using multimodal body sensing data. *7*, 99152-99160.

Gunes, H., & Piccardi, M. (2005). Affect recognition from face and body: early fusion vs. late fusion. 2005 IEEE international conference on systems, man and cybernetics,

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, *85*(1), 6-23.

Hall, D. L., & McMullen, S. A. (2004). *Mathematical techniques in multisensor data fusion*. Artech House.

Hallowell, M. R., & Gambatese, J. A. (2009). Activity-based safety risk quantification for concrete formwork construction. *Journal of Construction Engineering and Management*, *135*(10), 990-998.

Halpern, J. Y. (2017). *Reasoning about uncertainty*. MIT press.

Han, G., Choi, D., & Lim, W. (2007). A novel reference node selection algorithm based on trilateration for indoor sensor networks. 7th IEEE International Conference on Computer and Information Technology (CIT 2007),

Hassan, M. M., Uddin, M. Z., Mohamed, A., & Almogren, A. J. F. G. C. S. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *81*, 307-313.

He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. Proceedings of the IEEE conference on computer vision and pattern recognition,

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

He, Z., Cui, B., Zhou, W., & Yokoi, S. (2015). A proposal of interaction system between visitor and collection in museum hall by iBeacon. 2015 10th International Conference on Computer Science & Education (ICCSE),

Hutchinson, S. A., Cromwell, R. L., & Kak, A. C. (1988). Planning sensing strategies in a robot work cell with multi-sensor capabilities. Proceedings. 1988 IEEE International Conference on Robotics and Automation,

Huynh, T., & Schiele, B. (2005). Analyzing features for activity recognition. Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies,

Hwang, S., & Lee, S. (2017). Wristband-type wearable health devices to measure construction workers' physical demands. *Automation in Construction*, *83*, 330-340.

Ignatov, A. J. A. S. C. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *62*, 915-922.

Islam, M. T., Siddique, B. N. K., Rahman, S., & Jabid, T. (2018). Food image classification with convolutional neural network. 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS),

Jacobsen, E. L., Teizer, J., & Wandahl, S. (2023). Work estimation of construction workers for productivity monitoring using kinematic data and deep learning. *Automation in Construction*, *152*, 104932.

Jarkas, A. M. (2010). Critical investigation into the applicability of the learning curve theory to rebar fixing labor productivity. *Journal of Construction Engineering and Management*, *136*(12), 1279-1288.

Jeon, K. E., She, J., Soonsawad, P., & Ng, P. C. (2018). BLE Beacons for Internet of Things Applications: Survey, Challenges, and Opportunities. *Ieee Internet of Things Journal*, *5*(2), 811-828. https://doi.org/10.1109/Jiot.2017.2788449

Joshua, L., & Varghese, K. (2014). Automated recognition of construction labour activity using accelerometers in field situations. *International Journal of Productivity and Performance Management*, *63*(7), 841-862.

Kaempchen, N., Buehler, M., & Dietmayer, K. (2005). Feature-level fusion for free-form object tracking using laserscanner and video. IEEE Proceedings. Intelligent Vehicles Symposium, 2005.,

Kam, M., Zhu, X., & Kalata, P. (1997). Sensor fusion for mobile robot navigation. *Proceedings of the IEEE*, *85*(1), 108-119.

Kashimoto, Y., Morita, T., Fujimoto, M., Arakawa, Y., Suwa, H., & Yasumoto, K. (2017). Sensing activities and locations of senior citizens toward automatic daycare report generation. 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), Kendall, M. G. (1948). The advanced theory of statistics. Vols. 1. *The advanced theory of statistics. Vols. 1.*, *1*(Ed. 4).

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, *14*(1), 28-44.

Khoury, H. M., & Kamat, V. R. (2009). Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, *18*(4), 444-457.

Kim, B. K., Jung, H. M., Yoo, J.-B., Lee, W. Y., Park, C. Y., & Ko, Y. W. (2008). Design and implementation of cricket-based location tracking system. Proceedings of world academy of science, engineering and technology,

Kim, K., & Cho, Y. K. (2020). Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. *Automation in Construction*, *113*, 103126.

Kirstein, T. (2013). Multidisciplinary know-how for smart-textiles developers. Elsevier.

Kokar, M. M., Tomasik, J. A., & Weyman, J. (2004). Formalizing classes of information fusion systems. *Information Fusion*, *5*(3), 189-202.

Komai, K., Fujimoto, M., Arakawa, Y., Suwa, H., Kashimoto, Y., & Yasumoto, K. (2016). Beacon-based multi-person activity monitoring system for day care center. 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops),

Kong, L., Peng, X., Chen, Y., Wang, P., & Xu, M. (2020). Multi-sensor measurement and data fusion technology for manufacturing process monitoring: a literature review. *International journal of extreme manufacturing*, *2*(2), 022001.

Kotanen, A., Hannikainen, M., Leppakoski, H., & Hamalainen, T. D. (2003). Experiments on local positioning with Bluetooth. Proceedings ITCC 2003. International Conference on Information Technology: Coding and Computing,

Krishnan, N., Cook, D. J., & Wemlinger, Z. (2013). Learning a taxonomy of predefined and discovered activity patterns. *Journal of ambient intelligence and smart environments*, *5*(6), 621-637.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Kulkarni, S. C., & Rege, P. P. (2020). Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion*, *59*, 13-29.

Kuncheva, L. I. (2014). Combining pattern classifiers: methods and algorithms. John Wiley & Sons.

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, *12*(2), 74-82.

Kwon, M.-C., & Choi, S. (2018). Recognition of daily human activity using an artificial neural network and smartwatch. *Wireless Communications and Mobile Computing*, *2018*.

Laguna, J. O., Olaya, A. G., & Borrajo, D. (2011). A dynamic sliding window approach for activity recognition. International Conference on User Modeling, Adaptation, and Personalization,

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449-1477.

Lara, O. D., & Labrador, M. A. (2012). A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, *15*(3), 1192-1209.

Lee, S.-M., Yoon, S. M., & Cho, H. (2017). Human activity recognition from accelerometer data using Convolutional Neural Network. 2017 ieee international conference on big data and smart computing (bigcomp),

Lee, Y.-J., & Park, M.-W. (2019). 3D tracking of multiple onsite workers based on stereo vision. *Automation in Construction*, *98*, 146-159. https://doi.org/10.1016/j.autcon.2018.11.017

Li, H., Zhang, P., Al Moubayed, S., Patel, S. N., & Sample, A. P. (2016). Id-match: A hybrid computer vision and rfid system for recognizing individuals in groups. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,

Li, N., & Becerik-Gerber, B. (2011). Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment. *Advanced Engineering Informatics*, *25*(3), 535-546. https://doi.org/10.1016/j.aei.2011.02.004

Li, N., Becerik-Gerber, B., Krishnamachari, B., & Soibelman, L. (2014). A BIM centered indoor localization algorithm to support building fire emergency response operations. *Automation in Construction*, *42*, 78-89. https://doi.org/10.1016/j.autcon.2014.02.019

Li, N., Becerik-Gerber, B., Soibelman, L., & Krishnamachari, B. (2015). Comparative assessment of an indoor localization framework for building emergency response. *Automation in Construction*, *57*, 42-54. https://doi.org/10.1016/j.autcon.2015.04.004

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision,

Liu, D., Chen, J., Li, S., & Cui, W. (2018). An integrated visualization framework to support whole-process management of water pipeline safety. *Automation in Construction*, *89*, 24-37.

Liu, W., Wei, J., Liang, M., Cao, Y., & Hwang, I. (2013). Multi-sensor fusion and fault detection using hybrid estimation for air traffic surveillance. *IEEE Transactions on Aerospace and Electronic Systems*, *49*(4), 2323-2339.

Llinas, J., Bowman, C., Rogova, G., Steinberg, A., Waltz, E., & White, F. (2004). *Revisiting the JDL data fusion model II*.

Lu, M., Chen, W., Shen, X., Lam, H.-C., & Liu, J. (2007). Positioning and tracking construction vehicles in highly dense urban areas and building construction sites. *Automation in Construction*, *16*(5), 647-656. https://doi.org/10.1016/j.autcon.2006.11.001

Luo, R. C., & Kay, M. G. (1988). Multisensor integration and fusion: issues and approaches. Sensor Fusion,

Luo, R. C., Yih, C.-C., & Su, K. L. (2002). Multisensor fusion and integration: approaches, applications, and future research directions. *IEEE Sensors Journal*, *2*(2), 107-119.

Luo, X., O'Brien, W. J., & Julien, C. L. (2011). Comparative evaluation of Received Signal-Strength Index (RSSI) based indoor localization techniques for construction jobsites. *Advanced Engineering Informatics*, *25*(2), 355-363.

Ma, J. H., & Cha, S. H. (2020). A human data-driven interaction estimation using IoT sensors for workplace design. *Automation in Construction*, *119*, 103352. https://doi.org/10.1016/j.autcon.2020.103352

Ma, Z., Poslad, S., Bigham, J., Zhang, X., & Men, L. (2017). A BLE RSSI ranking based indoor positioning system for generic smartphones. 2017 Wireless Telecommunications Symposium (WTS),

Maalek, R., & Sadeghpour, F. (2016). Accuracy assessment of ultra-wide band technology in locating dynamic resources in indoor scenarios. *Automation in Construction*, *63*, 12-26.

Mackey, A., Spachos, P., & Plataniotis, K. N. (2018). Enhanced indoor navigation system with beacons and kalman filters. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Mackey, A., Spachos, P., Song, L., & Plataniotis, K. N. (2020). Improving BLE beacon proximity estimation accuracy through Bayesian filtering. *IEEE Internet of Things Journal*, *7*(4), 3160-3169.

Mahfouz, M. R., & Kuhn, M. J. (2011). UWB channel measurements and modeling for positioning and communications systems in the operating room. 2011 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems,

Malhotra, R. (1995). Temporal considerations in sensor management. Proceedings of the IEEE 1995 National Aerospace and Electronics Conference. NAECON 1995,

Martin, P., Ho, B.-J., Grupen, N., Muñoz, S., & Srivastava, M. (2014). An iBeacon primer for indoor localization: demo abstract. Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings,

Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, *57*, 115-129.

Minnen, D., Westeyn, T., Starner, T., Ward, J., & Lukowicz, P. (2006). Performance metrics and evaluation issues for continuous activity recognition. *Performance metrics for intelligent systems. NIST, Gaithersburg*, 141-148.

Montaser, A., & Moselhi, O. (2014). RFID indoor location identification for construction projects. *Automation in Construction*, *39*, 167-179.

Murphy, C. K. (2000). Combining belief functions when evidence conflicts. *Decision support systems*, *29*(1), 1-9.

Nasirzadeh, F., & Nojedehi, P. (2013). Dynamic modeling of labor productivity in construction projects. *International journal of project management*, *31*(6), 903-911.

Ng, P. C., She, J., & Ran, R. (2020). A reliable smart interaction with physical thing attached with ble beacon. *IEEE Internet of Things Journal*, *7*(4), 3650-3662.

Ng, S. T., & Tang, Z. (2010). Labour-intensive construction sub-contractors: Their critical success factors. *International Journal of Project Management*, *28*(7), 732-740.

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.

Obeidat, H., Shuaieb, W., Obeidat, O., & Abd-Alhameed, R. (2021). A review of indoor localization techniques and wireless technologies. *Wireless Personal Communications*, *119*(1), 289-327.

Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., & Bajcsy, R. (2013). Berkeley mhad: A comprehensive multimodal human action database. 2013 IEEE workshop on applications of computer vision (WACV),

Olson, E. (2010). A passive solution to the sensor synchronization problem. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems,

Palumbo, F., Barsocchi, P., Chessa, S., & Augusto, J. C. (2015). A stigmergic approach to indoor localization using bluetooth low energy beacons. 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS),

Pan, H., Liang, Z.-P., Anastasio, T. J., & Huang, T. S. (1998). A hybrid NN-Bayesian architecture for information fusion. Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269),

Panahandeh, G., Mohammadiha, N., Leijon, A., & Händel, P. (2013). Continuous hidden Markov model for pedestrian activity classification and gait analysis. *IEEE Transactions on Instrumentation and Measurement*, *62*(5), 1073-1083.

Papaioannou, S., Markham, A., & Trigoni, N. (2016). Tracking people in highly dynamic industrial environments. *IEEE Transactions on mobile computing*, *16*(8), 2351-2365.

Park, J., & Cho, Y. K. (2017). Development and Evaluation of a Probabilistic Local Search Algorithm for Complex Dynamic Indoor Construction Sites. *Journal of Computing in Civil Engineering*, *31*(4), 04017015. https://doi.org/10.1061/(asce)cp.1943-5487.0000658

Park, J., Kim, K., & Cho, Y. K. (2016). Framework of automated construction-safety monitoring using cloud-enabled BIM and BLE mobile tracking sensors. *Journal of construction engineering and management*, *143*(2), 05016019.

Park, J., Kim, K., & Cho, Y. K. (2017). Framework of automated construction-safety monitoring using cloud-enabled BIM and BLE mobile tracking sensors. *Journal of construction engineering and management*, *143*(2), 05016019.

Park, J., Marks, E., Cho, Y. K., & Suryanto, W. (2016). Performance test of wireless technologies for personnel and equipment proximity sensing in work zones. *Journal of Construction Engineering and Management*, *142*(1), 04015049.

Pelant, J., Tlamsa, Z., Benes, V., Polak, L., Kaller, O., Bolecek, L., Kufa, J., Sebesta, J., & Kratochvil, T. (2017). BLE device indoor localization based on RSS fingerprinting mapped by propagation modes. 2017 27th international conference radioelektronika (RADIOELEKTRONIKA),

Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors—a review of classification techniques. *Physiological measurement*, *30*(4), R1.

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, *80*, 241-265.

Qu, Y., Yang, M., Zhang, J., Xie, W., Qiang, B., & Chen, J. (2021). An outline of multi-sensor fusion methods for mobile agents indoor navigation. *Sensors*, *21*(5), 1605.

Rao, A. S., Radanovic, M., Liu, Y., Hu, S., Fang, Y., Khoshelham, K., Palaniswami, M., & Ngo, T. (2022). Real-time monitoring of construction sites: Sensors, methods, and applications. *Automation in Construction*, *136*, 104099.

Rednic, R., Gaura, E., Kemp, J., & Brusey, J. (2013). Fielded autonomous posture classification systems: design and realistic evaluation. 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing,

Rezazadeh, J., Subramanian, R., Sandrasegaran, K., Kong, X., Moradi, M., & Khodamoradi, F. (2018). Novel iBeacon Placement for Indoor Positioning in IoT. *IEEE Sensors Journal*, *18*(24), 10240-10247. https://doi.org/10.1109/jsen.2018.2875037

Ryu, J., Seo, J., Jebelli, H., & Lee, S. (2019). Automated action recognition using an accelerometer-embedded wristband-type activity tracker. *Journal of Construction Engineering and Management*, *145*(1), 04018114.

Sanhudo, L., Calvetti, D., Martins, J. P., Ramos, N. M., Meda, P., Goncalves, M. C., & Sousa, H. (2021). Activity classification using accelerometers and machine learning for complex construction worker activities. *Journal of Building Engineering*, *35*, 102001.

Sanhudo, L., Calvetti, D., Martins, J. P., Ramos, N. M., Mêda, P., Gonçalves, M. C., & Sousa, H. J. J. o. B. E. (2021). Activity classification using accelerometers and machine learning for complex construction worker activities. *35*, 102001.

Sarkar, S., Sarkar, S., Virani, N., Ray, A., & Yasar, M. (2014). Sensor fusion for fault detection and classification in distributed physical processes. *Frontiers in Robotics and AI*, *1*, 16.

Sathe, S., Papaioannou, T. G., Jeung, H., & Aberer, K. (2013). A survey of model-based sensor data acquisition and management. In *Managing and mining sensor data* (pp. 9-50). Springer.

Sebbak, F., & Benhammadi, F. (2017). Majority-consensus fusion approach for elderly IoT-based healthcare applications. *Annals of Telecommunications*, *72*(3-4), 157-171.

Sentz, K., & Ferson, S. (2002). Combination of evidence in Dempster-Shafer theory.

Seybold, J. S. (2005). *Introduction to RF propagation*. John Wiley & Sons.

Shafer, G. (1976). *A mathematical theory of evidence* (Vol. 42). Princeton university press.

Shafer, G. (1992). Dempster-shafer theory. *Encyclopedia of artificial intelligence*, *1*, 330-331.

Shao, H., Lin, J., Zhang, L., Galar, D., & Kumar, U. (2021). A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance. *Information Fusion*, *74*, 65-76.

Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *towards data science*, *6*(12), 310-316.

Shen, X., Cheng, W., & Lu, M. (2008). Wireless sensor networks for resources tracking at building construction sites. *Tsinghua Science and Technology*, *13*(S1), 78-83.

Sherafat, B., Ahn, C. R., Akhavian, R., Behzadan, A. H., Golparvar-Fard, M., Kim, H., Lee, Y.-C., Rashidi, A., & Azar, E. R. (2020). Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *Journal of construction engineering and management*, *146*(6), 03120002.

Sinha, A., Chen, H., Danu, D., Kirubarajan, T., & Farooq, M. (2008). Estimation and decision fusion: A survey. *Neurocomputing*, *71*(13-15), 2650-2656.

Smarandache, F., & Dezert, J. (2015). Advances and Applications of DSmT for Information Fusion. Collected Works, Volume 4.

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. *arXiv preprint arXiv:1505.00387*.

Steinberg, A. N., & Bowman, C. L. (2017). Revisions to the JDL data fusion model. In *Handbook of multisensor data fusion* (pp. 65-88). CRC press.

Steinberg, C. (1999). Bowman, and F. White:" Revisions to the JDL Data Fusion Model", in proc. Proceedings of SPIE, Sensor Fusion: Architectures, Algorithms, and Applications III,

Su, X., Tong, H., & Ji, P. (2014). Activity recognition with smartphone sensors. *Tsinghua science and technology*, *19*(3), 235-249.

Subhan, F., Hasbullah, H., Rozyyev, A., & Bakhsh, S. T. (2011). Indoor positioning in bluetooth networks using fingerprinting and lateration approach. 2011 International Conference on Information Science and Applications,

Teh, H. Y., Kempa-Liehr, A. W., & Wang, K. I.-K. (2020). Sensor data quality: A systematic review. *Journal of Big Data*, *7*(1), 1-49.

Teizer, J. (2015). Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. *Advanced Engineering Informatics*, *29*(2), 225-238.

Teizer, J., Neve, H., Li, H., Wandahl, S., König, J., Ochner, B., König, M., & Lerche, J. (2020). Construction resource efficiency improvement by Long Range Wide Area Network tracking and monitoring. *Automation in Construction*, *116*, 103245. https://doi.org/10.1016/j.autcon.2020.103245

Thomas, H. R., & Daily, J. (1983). Crew performance measurement via activity sampling. *Journal of Construction Engineering and Management*, *109*(3), 309-320.

Tubaishat, M., & Madria, S. (2003). Sensor networks: an overview. *IEEE potentials*, *22*(2), 20-23.

Tzirakis, P., Zafeiriou, S., & Schuller, B. (2019). Real-world automatic continuous affect recognition from audiovisual signals. In *Multimodal Behavior Analysis in the Wild* (pp. 387-406). Elsevier.

Urano, K., Kaji, K., Hiroi, K., & Kawaguchi, N. (2017). A location estimation method using mobile BLE tags with tandem scanners. Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers,

Vähä, P., Heikkilä, T., Kilpeläinen, P., Järviluoma, M., & Gambao, E. (2013). Extending automation of building construction—Survey on potential sensor technologies and robotic applications. *Automation in Construction*, *36*, 168-178.

Vakil, A., Liu, J., Zulch, P., Blasch, E., Ewing, R., & Li, J. (2021). A survey of multimodal sensor fusion for passive RF and EO information integration. *IEEE Aerospace and Electronic Systems Magazine*, *36*(7), 44-61.

Varsamou, M., & Antonakopoulos, T. (2014). A bluetooth smart analyzer in iBeacon networks. 2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin),

Vieira, S. T., Valadão, E., Rodríguez, D. Z., & Rosa, R. L. (2019). Wireless access point positioning optimization. 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM),

Vigneshwaran, S., Sen, S., Misra, A., Chakraborti, S., & Balan, R. K. (2015). Using infrastructure-provided context filters for efficient fine-grained activity sensing. 2015 IEEE international conference on pervasive computing and communications (PerCom),

Waltz, E., & Llinas, J. (1990). *Multisensor data fusion* (Vol. 685). Artech house Boston.

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018). Understanding convolution for semantic segmentation. 2018 IEEE winter conference on applications of computer vision (WACV),

Weiss, G. M., Timko, J. L., Gallagher, C. M., Yoneda, K., & Schreiber, A. J. (2016). Smartwatch-based activity recognition: A machine learning approach. 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI),

White, F. E. (1987). Data fusion lexicon, joint directors of laboratories, technical panel for C3, data fusion sub-panel. *San Diego, CA: Naval Ocean Systems Center*.

Wisanmongkol, J., Klinkusoom, L., Sanpechuda, T., Kovavisaruch, L.-o., & Kaemarungsi, K. (2019). Multipath mitigation for RSSI-based Bluetooth low energy localization. 2019 19th International Symposium on Communications and Information Technologies (ISCIT),

Wu, H., Siegel, M., Stiefelhagen, R., & Yang, J. (2002). Sensor fusion using Dempster-Shafer theory [for context-aware HCI]. IMTC/2002. Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276),

Xu, C., Chai, D., He, J., Zhang, X., & Duan, S. J. I. A. (2019). InnoHAR: A deep neural network for complex human activity recognition. *7*, 9893-9902.

Xu, P., Davoine, F., Bordes, J.-B., Zhao, H., & Denœux, T. (2016). Multimodal information fusion for urban scene understanding. *Machine Vision and Applications*, *27*(3), 331-349.

Xu, S., Wang, Y., Sun, M., Si, M., & Cao, H. (2021). A Real-Time BLE/PDR Integrated System by Using an Improved Robust Filter for Indoor Position. *Applied Sciences*, *11*(17), 8170.

Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. *Information sciences*, *41*(2), 93-137.

Yang, J., Arif, O., Vela, P. A., Teizer, J., & Shi, Z. (2010). Tracking multiple workers on construction sites using video cameras. *Advanced Engineering Informatics*, *24*(4), 428-434.

Yao, R., Lin, G., Shi, Q., & Ranasinghe, D. C. (2018). Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. *Pattern Recognition*, *78*, 252-266.

Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, *21*(6), 2140.

Yi, W., & Chan, A. P. (2014). Critical review of labor productivity research in construction journals. *Journal of management in engineering*, *30*(2), 214-225.

Zafari, F., Papapanagiotou, I., Devetsikiotis, M., & Hacker, T. (2017). An ibeacon based proximity and indoor localization system. *arXiv preprint arXiv:1703.07876*.

Zappi, P., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., & Troster, G. (2007). Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. 2007 3rd international conference on intelligent sensors, sensor networks and information,

Zhang, H., Zhou, W., & Parker, L. E. (2014). Fuzzy segmentation and recognition of continuous human activities. 2014 IEEE International Conference on Robotics and Automation (ICRA),

Zhang, M., Cao, T., & Zhao, X. (2017). Applying sensor-based technology to improve construction safety management. *Sensors*, *17*(8), 1841.

Zhang, M., Shi, R., & Yang, Z. (2020). A critical review of vision-based occupational health and safety monitoring of construction site workers. *Safety science*, *126*, 104658.

Zhang, S., McCullagh, P., Nugent, C., & Zheng, H. (2010). Activity monitoring using a smart phone's accelerometer with hierarchical classification. 2010 sixth international conference on intelligent environments,

Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2021). Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, *129*(11), 3069-3087.

Zhao, J., Seppänen, O., Peltokorpi, A., Badihi, B., & Olivieri, H. (2019). Real-time resource tracking for analyzing value-adding time in construction. *Automation in Construction*, *104*, 52-65.

Zheng, X., Wang, M., & Ordieres-Meré, J. (2018). Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0. *Sensors*, *18*(7), 2146.