Integrated Mobile Sensor-Based Activity Recognition of Construction Equipment and Human Crews

Reza Akhavian
Lianne Brito
Amir Behzadan
Department of Civil, Environmental, and Construction Engineering
University of Central Florida
12800 Pegasus Dr.
Orlando, FL 32816
reza@knights.ucf.edu
lianne.brito@ucf.edu
amir.behzadan@ucf.edu

ABSTRACT

Automated activity recognition of heavy construction equipment as well as human crews can contribute to correct and accurate measurement of a variety of construction and infrastructure project performance indicators. Productivity assessment through work sampling, safety and health monitoring using worker ergonomic analysis, and sustainability measurement through equipment activity cycle monitoring to eliminate ineffective and idle times thus reducing greenhouse gas emission (GHG), are some potential areas that can benefit from the integration of automated activity recognition and analysis techniques. Despite their proven performance and applications in other domains, few construction engineering and management (CEM) studies have so far employed non-vision sensing technologies for construction equipment and workers’ activity recognition. The existence of a variety of sensors in ubiquitous smartphones with evolving computing, networking, and storage capabilities has created great opportunities for a variety of pervasive computing applications. In light of this, this paper describes the latest findings of an ongoing project that aims to design and validate a ubiquitous smartphone-based automated activity recognition framework using built-in accelerometer and gyroscope sensors. Collected data are segmented to overlapping windows to extract time- and frequency domain features. Since each sensor collected data in three axes (x, y, z), several features from all three axes are extracted to ensure device placement orientation independency. Finally, features are used as the input of supervised machine learning classifiers. The results of the experiments indicate that the trained models are able to classify construction workers and equipment activities with over 90% overall accuracy.

Key words: activity recognition—big data analytics—construction engineering—machine learning—mobile sensor

INTRODUCTION

With the ever-growing infrastructure projects, demands for information and automation technologies in the architecture, engineering, construction, and facility management (AEC/FM) industry is rapidly increasing. 3D printing in design and construction of affordable houses (Krassenstein 2015), drones for various applications in construction jobsites (Irizarry et al. 2012), and Micro-Electro-Mechanical Systems (MEMS) sensors for tracking and monitoring of construction resources (Akhavian and Behzadan 2015) are some of the examples of how latest
cutting-edge technology serves the AEC/FM industry. The latter, in particular, has ample unrivalled potentials for use in day-to-day operations due to the existence of various sensors in a major technology platform carried by almost everyone these days, the smartphone.

Ubiquity, affordability, small size, and computing power of mobile phones, equipped with a host of sensors have made them ideal choices for tracking and monitoring of construction resources. In particular, these built-in sensors can provide invaluable information regarding the performance, safety, and behavior of construction workers and equipment in the field. For example, activity analysis using inertial measurement unit (IMU) sensors including accelerometer and gyroscope can help evaluate the time spent on interconnected construction tasks. Such information results in better understanding and potential improvement of the processes involved. Moreover, effective and timely analysis and tracking of the construction resources can help in productivity measurement, progress evaluation, labor training programs, safety and health management, and greenhouse gas emission (GHG) and fuel consumption analysis in construction projects.

This paper presents the results of a smartphone sensors-based machine learning platform for accurate recognition and classification of activities performed by construction equipment and workers. Specifically, process data is collected using smartphone built-in accelerometer and gyroscope from construction equipment and human crews. Certain data pre-processing is then performed to prepare the data as the training input of supervised machine learning classifiers. The outputs of the classifiers are the labels of various activities carried out by the construction resources and detected using this framework.

LITERATURE REVIEW

Human Activity Recognition using Sensors

The initial efforts on human activity recognition date back to the late '90s (Foerster et al. 1999). During the last 15-16 years, the use of MEMS sensors has increased tremendously for acquiring knowledge on humans' activities. With the growing demand in activity recognition for different fields of research and practice, the cost of system implementation and prototyping has also decreased, which makes the technology more affordable and accessible to businesses and the industry. Most of the activity recognition research during the past decade has been conducted using wired or wireless accelerometers attached to different parts of a subject’s body. It has been proven that the accuracy of the recognition algorithm improves as sensors are attached to more than one part of the body, while a disadvantage of this approach is that the presence of multiple asynchronous sensors each communicating through a separate interface make the data collection process computationally inefficient and tedious, while ergonomically obtrusive and uncomfortable for the subject. In a study by Frank et al. (2010) MEMS-based IMUs attached to the performer’s waist were used to present four different algorithms and recognize daily activities such as standing, running, walking, and falling. Even though most of the activities were recognized with an accuracy of 93-100%, the rate of accuracy of detecting falling was 80%.

Activity recognition has been widely used in the medical field primarily for elderly patient monitoring. In a study by Gupta and Dallas (2014), waist-mounted triaxial accelerometer was used to classify gait events into six daily living activities including run, jump, walk, sit and transitional events including sit to stand and stand to kneel. While most of the previous studies measured the activities with accelerometer sensor, some studies added gyroscope sensor to improve the accuracy of classification. For instance, in a study aiming at accurate and fast fall
detection, Li et al. (2009) proved that using both accelerometer and gyroscope data for activity recognition yields in more accurate results than using accelerometer only. They reported specific improvement in classifying dynamic transitions rather than static postures and in determining whether the transitional movement was intentional. In another study, Frank et al. (2010) used MEMS-based IMU to train four different classification algorithms that recognized daily activities such as standing, sitting, walking, running, jumping, falling, and lying. In these and most other similar studies, the target activities for classification are daily routine tasks. However, activities in a construction jobsite are relatively more complex and involve interactions between multiple resources (human and equipment crews), which highlights the need for rigorous research in this area within the AEC/FM domain.

Design and development of a small-size, unobtrusive, and low-cost data collection scheme has been one of the most important challenges in improving the accuracy of activity recognition in a more affordable and robust manner (Lara and Labrador 2013). During the last decade and with the advancement of smartphones, a major paradigm shift occurred in the data collection scheme for human activity recognition. As mentioned before, nowadays smartphones include a wide variety of computationally powerful sensors. Many researchers seized this opportunity to build more pervasive human activity recognition systems. Kwapisz et al. (2011) conducted a study on 29 Android smartphone users that utilized accelerometer sensor to monitor daily activities such as walking, jogging, ascending stairs, descending stairs, sitting, and standing for a period of time. The most accurate classification results were achieved in two relatively straightforward activities: sitting and jogging (compared to walking, standing, and climbing up/down the stairs). Again, even though the target activities were recognized with 90% accuracy in most cases (except for climbing up/down the stairs), recognizing such daily and straightforward activities is relatively easier than those carried out in a construction site, for instance. Dernbach et al. (2012) published the results of their study aimed at measuring simple and more complex daily activities using built-in accelerometer and gyroscope sensors in smartphones. Simple activities included biking, climbing stairs, driving, lying, running, sitting, standing, and walking while complex activities included cleaning (i.e. wiping down the kitchen counter top and sink), cooking (i.e. heating a bowl of water in the microwave and pouring a glass of water from a pitcher in the fridge), medication (i.e. retrieving pills from the cupboard and sorted out a week’s worth of doses), sweeping (i.e. sweeping the kitchen area), washing hands, and watering plants. Their results showed that while an accuracy of around 90% was achieved for simple activities, more complex activities were best classified with less than 50% accuracy. Moreover, in all activity types, adding gyroscope data to accelerometer data improved the results by 10-12%. Another research that used smartphone accelerometer and gyroscope sensors was conducted by Wu et al. (2012). In this research an iPod Touch was used for data collection measuring 13 different daily activities (e.g. stair climbing, jogging, sitting) from 16 human subjects. Using only the accelerometer data, the accuracy in recognizing these activities ranged between 50% to 100%, while adding the gyroscope data helped improve the results by another 3.1% to 13.4%.

Non-Human Subject Activity Recognition using Sensors

In addition to human activity recognition, classification of activities performed by construction equipment is investigated in this research. To the best of the authors’ knowledge, very little research has been conducted aiming at non-human subject activity recognition since the application areas are rare and limited. For example, in transportation research and practice, the use of sensors for transportation mode detection, urban planning improvement, targeted advertising, and guidance systems has recently gained traction (Hemminki et al. 2013; Wang et al. 2010; Zheng et al. 2010). Although most of these studies primarily relied on global
positioning data (GPS) data for recognizing activities and in particular, transportation modes, some of them have used accelerometer data for more precision. In one study, Reddy et al. (2010) used GPS and accelerometer to classify stationary, walking, running, biking and motorized transportation. In their work, classification mainly relied on GPS speed for detecting motorized mode and no distinction was made between different motorized modalities. However, in another study aiming at improving this framework, Hemminki et al. (2013) used only accelerometer data to detect if the subject is stationary or on a motorized transportation and to classify the motorized transportation modalities into bus, train, metro, tram, or car. Using some improved algorithms for estimating the gravity component of the accelerometer measurements and feature extraction, they achieved higher precision and recall than the combined GPS and accelerometers approach while eliminating the high power consumption of the GPS sensor. In addition to transportation systems management and planning, a few construction engineering and management (CEM) research studies targeted non-human subject activity recognition that are summarized in the next Subsection.

Activity Recognition in Construction

Considering the dynamic and complex environment of most construction project sites, being able to control and measure the efficiency of construction resources is vital to the overall performance of the project in terms of time and financial resources. Moreover, by monitoring workers and equipment activities, catastrophes that include safety and health issues as well as many lawsuits could be prevented. Within the CEM research domain, vision-based systems are used in most existing object recognition and tracking frameworks. For example, Brilakis et al. (2011) proposed a framework for vision-based tracking of construction entities. Their methodology requires calibration of two cameras, recognition of construction resources and identification of the corresponding regions, matching the entities identified in different cameras, two-dimensional (2D) tracking of the matched entities, and finally calculating the 3D coordinates. This and similar vision-based approaches, although provide promising results for recognition and tracking of construction equipment, still require much computation in each one of the aforementioned steps. In another study, an image processing methodology was adopted for idle time quantification of hydraulic excavators (Zou and Kim 2007). The level of detail (LoD) of the framework, however, was limited to detection of only idle and busy states of a hydraulic excavator. For the purpose of learning and classification of labor and equipment actions, the concept of Bag-of-Video-Feature-Words model was extended into the construction domain (Gong et al. 2011). This technique uses unsupervised learning for classification, and only considers frequency of feature occurrence for classification. Another vision-based framework was proposed by Rezazadeh Azar and McCabe (2012) for dirt-loading cycles in earthmoving operations that depends on the location of equipment which requires the algorithm to be modified for every new jobsite.

Construction workers’ activity recognition using vision-based systems has been also the subject of some other studies. For example, 3D range image cameras were used for tracking and surveillance of construction workers for safety and health monitoring (Gonsalves and Teizer 2009; Peddi et al. 2009). Gonsalves and Teizer (2009) indicated that if their proposed system is used in conjunction with artificial neural network (ANN), results would be more robust for prevention of fatal accidents and related health issues. In their study on construction workers’ unsafe actions, Han and Lee (2013) developed a framework for 3D human skeleton extraction from video to detect unsafe predefined motion templates. All of these frameworks, although presented successful results in their target domain, require installation of multiple cameras (up to 8 in some cases), have short recognition distance (maximum of 4 meters for Kinect) and
require direct line of sight for implementation. Such shortcomings have served as a major motivation to investigate alternative solutions that can potentially alleviate these problems.

Non-vision-based (a.k.a. sensor-based) worker activity analysis has recently gained popularity among CEM researchers. Cheng et al. (2013) used ultra-wide band (UWB) and Physiological Status Monitors (PSMs) for productivity assessment. However, the LoD in recognizing the activities was limited to identification of traveling, working, and idling states of workers and could not provide further insight into identified activities. In another set of research studies aiming at construction equipment activity analysis to support process visualization, remote monitoring and planning, queueing analysis, and knowledge-based simulation input modeling, the authors developed a framework by fusing data from ultra-wide band (UWB), payload, and orientation (angle) sensors to build a spatio-temporal taxonomy-based reasoning scheme for activity classification in heavy construction (Akhavian and Behzadan 2012, 2013, 2014).

As one of the first accelerometer-based activity recognition studies, Joshua and Varghese (2011) developed a work sampling framework for bricklayers. The scope of that study, however, was limited to only a single bricklayer in a controlled environment. Moreover, their proposed framework used accelerometer as the sole source of motion data. Also, the necessity of installing wired sensors on the worker’s body may introduce a constraint on the worker’s freedom of movement. In another study, Ahn et al. (2013) used accelerometers to classify an excavator operations into three modes of engine-off, idling, and working. Further decomposition of these activities, however, was not explored in their study.

**METHODOLOGY**

In this study, data are collected using mobile phone accelerometer and gyroscope sensors. Collected raw sensory data are segmented into windows containing certain number of data points. Next, key statistical features are calculated within each window. Furthermore, each segment is labeled based on the corresponding activity class performed at the time identified by the timestamp of the collected data. In order to train a predictive model, five classifiers of different types are used to recognize activities performed in the data collection experiments. Figure 1 depicts the steps from data collection to activity recognition.

![Figure 1. System architecture for activity recognition of construction resources using mobile phones](image-url)
Data Collection

Data collection is performed using commercially available data logger applications for iOS and/or Android devices. Accelerometer sensors measure the acceleration of the device. The reading can be in one, two, or all three axes of x, y, and z. The raw data is represented as a set of vectors and returned together with a timestamp of the reading. Gyroscope is a sensor that measures the rotation rate of the device by detecting the roll, pitch, and yaw motions of the smartphone about the x, y, and z axes. Similar to accelerometer, readings are presented as time-stamped vectors. When the mobile device is attached to construction equipment or worker’s body involved in different activities, these two sensors generate different (and distinct) patterns in their transmitted signals.

Data Preparation

A major step before transforming raw data into the input features for machine learning algorithms is removing noise and accounting for missing data. When collecting data for a long period of time, it is possible that the sensors temporarily freeze or fail to properly collect and store data for fractions of a second to a few seconds and in return, compensate for the missing data points by collecting data in a rate higher than the assigned frequency. In such cases, a preprocessing technique to fill in for missing data points and removing redundant ones can help insuring a continuous and orderly dataset. Also, since the raw data are often collected with a high sampling rate, segmentation of the data helps in data compression and prepares data for feature extraction (Khan et al. 2011). If segmentation is performed considering an overlap between adjacent windows, it reduces the error caused by the transition state noise (Su et al. 2014). The length of the window size depends on the sampling frequency and the nature of activities targeted for classification from which data is collected (Su et al. 2014).

Feature Extraction

Feature is an attribute of the raw data that should be calculated (Khan et al. 2011). In data analytics applications, statistical time- and frequency-domain features generated in each window are used as the input of the training process (Ravi et al. 2005). The ability to extract appropriate features depends on the application domain and can steer the process of retaining the relevant information. Most previous studies on activity recognition used almost the same set of features for training the models and classification of activities (Shoaib et al. 2015).

Data Annotation

Following data segmentation and feature extraction, the corresponding activity class labels should be assigned to each window. This serves as the ground truth for the learning algorithm and can be retrieved from the video recorded at the time of the experiment.

Supervised Learning

In supervised learning classification, class labels discussed in Data Annotation (above) are provided to the learning algorithms to generate a model or function that matches the input (i.e. features) to the output (i.e. activity classes) (Ravi et al. 2005). The goal is to infer a function using examples for which the class labels are known (i.e. training data). The performance of this function is evaluated by measuring the accuracy in predicting the class labels of unseen examples. Researchers have used different types of supervised classification methods for activity recognition (Kim et al. 2013; Reddy et al. 2010; Sun et al. 2010). Details of the supervised learning classification algorithms are discussed in the next Section.
Model Assessment
In order to determine the reliability of the trained model in detecting new examples of activity classes, part of the training dataset is used for testing the model. It is recommended that the test set is independent of the training set, meaning that the data that are used for testing have not been among the training data. For example, randomly chosen 10% of the training data can be left out so that the training is performed on the remaining 90% of the data. Assessment of the model provides an opportunity for its fine-tuning so that certain variables (e.g. regularization factor to prevent over-fitting in ANNs) in the algorithm can be revised to yield the best possible model.

Activity Recognition
Once the model is trained and its parameters are finalized, it can be used for recognizing activities for which it has been trained. While data is being collected to determine the activities according to a trained classifier, such data can be stored in a dataset repository and be added to the existing training data, so that the model is further trained with a richer training dataset.

DETAILS OF THE CLASSIFICATION ALGORITHMS
In the presented research, five different classification techniques are used in order to systematically evaluate their performance in accurately detecting construction activities. In particular, ANN, decision tree, K-nearest neighbor (KNN), logistic regression, and support vector machine (SVM) are employed for classification. Decision tree, KNN, and SVM have been previously used for activity recognition (Kose et al. 2012; Ravi et al. 2005; Yan et al. 2012) so they are selected in this study to evaluate their performance for classifying construction activities. However, ANN and logistic regression were examined to a much lesser extent (Staudenmayer et al. 2009).

Artificial Neural Network (ANN)
An ANN trained based on the experiments’ data follows a simple pattern of one input, one hidden, and one output layer. The number of input layer units is set to the number of extracted features. The hidden layer consists of $p=25$ units. The number of units for the output layer is equal to the number of activity classes, $n$ in each case. Considering the large feature space and in order to prevent over-fitting, regularization was used. Using a regularization parameter, the magnitude of the model weights decreases, so that the model will not suffer from high variance to fail to generalize to the new unseen examples (Haykin et al. 2009). The activation function (i.e. hypothesis) used for minimizing the cost function in the training process is a Sigmoid function shown in Equation 1,

$$h_{\Phi}(x) = \frac{1}{1+e^{-\Phi x}}$$  \hspace{1cm} (1)

in which $h(x)$ is the activation function (i.e. hypothesis), $\Phi$ is a matrix of model weights (i.e. parameters), and $x$ is the features matrix. In this study, in order to minimize the cost function, the most commonly used ANN training method, namely feed-forward backpropagation is used. Considering a set of randomly chosen initial weights, the backpropagation algorithm calculates the error of the activation function in detecting the true classes and tries to minimize this error by taking subsequent partial derivatives of the cost function with respect to the model weights (Hassoun 1995).
**Decision Tree**

Decision tree is one of the most powerful yet simplest algorithms for classification (Bishop 2006). The decision tree method that is used in this research is classification and regression tree (CART). CART partitions the training examples in the feature space into rectangle regions (a.k.a. nodes) and assigns each class to a region. The process begins with all classes spread over the feature space and examines all possible binary splits on every feature (Bishop 2006). A split is selected if it has the best optimization criterion which is the Gini diversity index in this research, as shown in Equation 2,

\[ l_G(f) = 1 - \sum_{i=1}^{k} f_i^2 \]  
\[ (2) \]

in which \( l_G \) is the Gini index, \( f_i \) is the fraction of items labeled with value \( i \) and \( k \) is the number of classes. The process of splitting is repeated iteratively for all nodes until they are pure. A node is considered pure if it contains only observations of one class, implying a Gini index of zero, or that there are fewer than 10 observations to split.

**K-Nearest Neighbor (KNN)**

Similar to the decision tree and unlike the ANN, KNN is a simple algorithm. Training examples identified by their labels are spread over the feature space. A new example is assigned to a class that is most common amongst its K nearest examples considering the Euclidean distance that is used as the metric in this research, and as appears in Equation 3,

\[ D = \sqrt{(x_i^{(1)} - x_{new}^{(1)})^2 + (x_i^{(2)} - x_{new}^{(2)})^2 + \cdots + (x_i^{(d)} - x_{new}^{(d)})^2} \]  
\[ (3) \]

in which \( D \) is the Euclidean distance, \( x_i \) is an existing example data point which has the least distance with the new example, \( x_{new} \) is the new example to be classified, and \( d \) is the dimension of the feature space.

**Logistic Regression**

Logistic regression is a type of regression problems in which the output is discretized for classification (Hastie et al. 2009). Logistic regression seeks to form a hypothesis function that maps the input (i.e. training data) to the output (i.e. class labels) by estimating the conditional probability of an example belonging to class \( k \) given that the example actually belongs to the class \( k \). This is accomplished by minimizing a cost function using a hypothesis function and correct classes to find the parameters of the mapping model (Hastie et al. 2009). The hypothesis function used in this research is the same as the activation function introduced in Equation 1 (the Sigmoid function) and thus the cost function to minimize is as shown in Equation 4,

\[ J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^{m} y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right] \]  
\[ (4) \]

in which \( J(\theta) \) is the cost function, \( m \) is the number of training examples, \( x^{(i)} \) is the \( i \)th training example, and \( y^{(i)} \) is the corresponding correct label. Once the cost function is minimized using any mathematical method such as the Gradient Decent (Hastie et al. 2009) and parameters are found, the hypothesis will be formed. In multi-class classification, the one-versus-all method is
used to determine if a new example belongs to the class $k$ (Hastie et al. 2009). Therefore, considering $k$ classes, $k$ hypothesis functions will be evaluated for each new example and the one that results in the maximum hypothesis is selected.

**Support Vector Machine (SVM)**

Compared to decision tree and KNN, SVM is considered as a more powerful classification algorithm. Although it has been widely used in vision-based pattern recognition and classification problems, some researchers (Bishop 2006) used it for classifying daily activities and thus its performance is also assessed in this research. In a nutshell, SVM tries to maximize the margin around hyperplanes that separate different classes from each other. SVM can benefit from a maximum margin hyperplane in a transformed feature space using kernel function to create non-linear classifiers. The kernel function used for non-linear classification in this research is Gaussian radial basis function (rbf) which has been successfully applied in the past to activity recognition problems (Ravi et al. 2005). Further description of SVM models are out of the scope of this study but can be found in (Bishop 2006).

**VALIDATION EXPERIMENTS**

In this Section, the description and details of two separate experiments conducted using construction equipment and workers in order to validate the designed activity recognition methods are provided.

**Front-End Loader Activity Recognition**

In this experiment, smartphones were placed inside the equipment cabin for data collection. At any time, two smartphones were simultaneously used to guarantee the uninterrupted storage of data. It must be noted that since data collection and feature extraction is done using tri-axial data, results do not depend on the placement orientation of the data collection device. Moreover, potential significant correlation between each pair of axes is reflected in three of the extracted features, thus guaranteeing capturing any distinguishable feature related to the placement orientation of the data collection devices. In order to fully automate the process of data collection, low-cost near field communication (NFC) RFID smart tags were also used (Want 2006). NFC tags were glued to the device holder (i.e. suction cup attached to the side window of the cabin) to automatically launch the data logger application once the smartphone was placed in the holder. A JOHN DEERE 744J front-end loader was employed for data collection. All experiment operations were fully videotaped for later activity annotation and labeling, and visual validation. Figure 2 shows how data collection devices were mounted and secured inside the target equipment cabin.

![Figure 2. Smartphones mounted inside the front-end loader cabin](image)
Data Collection

Data was collected using commercially available data logger applications for iOS and Android devices. The sampling frequency was set at 100 Hz. Among different modes of data collected in this study, it was observed that acceleration (i.e., vibration) values resulted from different equipment motions had the highest degree of volatility. Several sensor manufacturers have recommended that a bandwidth of 50 Hz be used for normal-speed vibration and tilt sensing applications. Therefore, in this research, and considering the Nyquist criterion in signal processing (Lyons et al. 2005), the sampling frequency was set at twice this value or 100 Hz. This bandwidth guaranteed that no significant motion was overlooked and at the same time, the volume of recorded data was not prohibitively large. Data was stored with comma separated value (CSV) format for processing in Microsoft Excel. The logger applications provided time-stamped data which facilitated the synchronization of data and video recordings. As mentioned earlier, GPS data was not directly used in data mining processes employed in this study and was only collected to demonstrate the potential of acquiring high accuracy positional data for such context-aware applications. Figure 3 shows snapshots of the collected accelerometer, gyroscope, and GPS data.

Data Processing and Classification

Raw data must be first represented in terms of specific features over a window of certain data points. In this research, mean, variance, peak, interquartile range (IQR), correlation, and root
mean error (RMS) are the statistical time-domain features that were extracted from data. Moreover, signal energy was picked as the only frequency-domain feature since it had already shown positive discrimination results in previous studies (Figo et al. 2010; Khan et al. 2011) for context recognition using accelerometer data. These 7 features were extracted from both accelerometer and gyroscope data corresponding to each of the x, y, and z axis. Since both sensors return tri-axial values (x, y, z), a total of 42 (i.e. multiplication of 7 features from 2 sensors in 3 axes) features were extracted. The size of the window depends on the sampling frequency and thus, varies for different applications. However, it should be selected in such a way that no important action is missed. This can be achieved by overlapping consecutive windows. Previous studies using accelerometer for context recognition have suggested a 50% overlap between windows (Ahn et al. 2013; Darren Graham et al. 2005; DeVaul and Dunn 2001). Time-domain features can be extracted using statistical analysis. However, the frequency-domain feature (i.e. signal energy) should be extracted from the frequency spectrum which requires signal transformation. In this study, fast Fourier transform (FFT) was used to convert the time-domain signal to the frequency-domain. In order to be computationally efficient, FFT requires the number of data points in a window to be a power of 2. Data was initially segmented into windows of 128 data points with 50% overlap. Therefore, given a sampling frequency of 100 Hz, each window contained 1.28 seconds of the experiment data. A sensitivity analysis presented in Section 4.5 provides more detail about the process of selecting the proper window size. The entire data analysis process including feature extraction was performed in Matlab.

Among all extracted features, there are some that may not add to the accuracy of the classification. This might be due to the correlation that exists among the collected data and consequently extracted features, since many actions result in a similar pattern in different directions and/or different sensor types (i.e. accelerometer vs. gyroscope). Therefore, in order to reduce the computational cost and time of the classification process, and increase its accuracy, a subset of the discriminative features is selected by filtering out (removing) irrelevant or redundant features (Pirttikangas et al. 2006). In this study, two filtering approaches are used: ReliefF and Correlation-based Feature Selection (CFS). ReliefF is a weighting algorithm that assigns a weight to each feature and ranks them according to how well their values distinguish between the instances of the same and different classes that are near each other (Yu and Liu 2003). CFS is a subset search algorithm that applies a correlation measure to assess the goodness of feature subsets based on the selected features that are highly correlated to the class, yet uncorrelated to each other (Hall 1999).

Using CFS, irrelevant and redundant features were removed which yielded 12 features (out of 42). These features were then ranked by ReliefF using their weight factors. The first 12 features selected by ReliefF were compared to those selected by CFS and the 7 common features in both methods were ultimately chosen as the final feature space. Table 1 shows the selected features by each filter as well as their intersection.

A learning algorithm can be supervised or unsupervised depending on whether or not different classes are labeled for training. Although unsupervised methods can be employed for equipment action recognition (Gong et al. 2011), supervised learning algorithms provide better performance for this purpose (Golparvar-Fard et al. 2013). This is mainly due to the fact that action classes of a piece of equipment consist of some classes with limited number of instances. This creates an imbalanced set of classes (caused by large differences between the number of instances in some classes) that can very likely lead to over-fitting in unsupervised
learning classification. Among several supervised learning methods those that follow more complex algorithms may seem more accurate in classification.

### Table 1. Selected features by CFS and ReliefF and their intersection (A: Accelerometer, G: Gyroscope)

<table>
<thead>
<tr>
<th>Filter</th>
<th>Selected Features</th>
<th>Common Selected Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CFS</strong></td>
<td>A_mean_x, A_mean_y, A_mean_z, A_peak_x, A_iqr_y, A_iqr_z, A_correlation_z, A_rms_z, G_mean_x, G_mean_y, G_mean_z, G_variance_x</td>
<td>G_mean_z, A_mean_x, G_mean_x, A_mean_y, A_mean_z, A_iqr_y, A_iqr_z, A_correlation_z, A_rms_z</td>
</tr>
<tr>
<td><strong>ReliefF</strong></td>
<td>G_mean_z, A_mean_x, G_mean_x, A_peak_z, A_mean_y, A_correlation_y, A_correlation_x, A_mean_z, A_iqr_z, A_peak_x, A_peak_y, G_rms_z</td>
<td>A_peak_x</td>
</tr>
</tbody>
</table>

However, the choice of the learning algorithm is highly dependent on the characteristics and volume of data. As a result, a “single” best classifier does not generally exist and each case requires unique evaluation of the learning algorithm through cross validation (Goldberg 1989). Therefore, a number of learning algorithms are tested in this research to compare their performance in classifying actions using sensory data.

In this experiment, classification was performed by labeling the classes in different LoDs. The first set of training and classification algorithms is applied to three classes namely Engine Off, Idle, and Busy. Next, the Busy class is broken down into two subclasses of Moving and Scooping, and Moving and Dumping, and so on. As stated earlier, for action classification, five supervised learning methods were used: 1) Logistic Regression, 2) K-NN, 3) Decision Tree, 4) ANN (feed-forward backpropagation), and 5) SVM. Using different classifiers reduces the uncertainty of the results that might be related to the classification algorithm that each classifier uses.

### Results of Equipment Activity Recognition

For each LoD, five classifiers were trained. Training and testing were performed through stratified 10-fold cross validations. In a \(k\)-fold cross validation the dataset is divided into \(k\) sets of equal sizes, and classifiers are trained \(k\) times, each time they are tested on one of the \(k\) folds and trained using the remaining \(k \sim 1\) folds. Moreover, in the stratified \(k\)-fold cross validation, each fold contains almost the same proportions of classes as in the whole dataset. The mean accuracy is reported as the accuracy of each class. Result of the classification performance for each case (i.e. LoD) is presented in Table 2 in terms of overall classifier accuracy. As shown in Table 2, ANNs had the best relative overall accuracy among all five classifiers in all the LoDs. Moreover, although the 3-class level the accuracy gets to as high as 98.59%, the highest accuracy for the 4-class level is 81.30% which is less than that of the 5-class level, which is 86.09%.
Table 2. Overall accuracy of classifiers for each LoD

<table>
<thead>
<tr>
<th></th>
<th>Classifier</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Classes</td>
<td>ANN</td>
<td>98.59</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>97.40</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>97.65</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>96.93</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>96.71</td>
</tr>
<tr>
<td>4 Classes</td>
<td>ANN</td>
<td>81.30</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>81.21</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>80.51</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>77.58</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>78.03</td>
</tr>
<tr>
<td>5 Classes</td>
<td>ANN</td>
<td>86.09</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>73.78</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>84.20</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>84.42</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>78.58</td>
</tr>
</tbody>
</table>

As stated earlier, construction equipment activity recognition has been previously explored through vision-based technologies. Gong et al. (2011) reported an overall accuracy of 86.33% for classification of three action classes of a backhoe. In a more recent study, Golparvar-Fard et al. (2013) achieved 86.33% and 76.0% average accuracy for three and four action classes of an excavator, respectively, and 98.33% average accuracy for three action classes of a dump truck. Although the target construction equipment are different in each case and action categories varies in these studies, the developed framework in this study that uses IMUs for the first time for construction equipment action recognition shows promising results when compared to existing vision-based systems that have been the subject of many research studies in the past few years.

Construction Workers’ Activity Recognition

In a separate set of experiments, an outdoor construction workspace was created where different activities were performed by multiple workers. These activities included sawing, hammering, turning a wrench, loading sections into wheelbarrows, pushing loaded wheelbarrows, dumping sections from wheelbarrows, and returning with empty wheelbarrows. Activities were performed in 3 different categories in order to assess certain circumstances (as described later) in the outcome of classification. A Commercially available armband was used to secure a smartphone on the upper arm of the dominant hand of each worker. Recent research on the selection of accelerometer location on bricklayer’s body for activity recognition has shown that according to the body movement of the worker while performing different bricklaying activities, among 15 potential locations for wearing an accelerometer, the lower left arm and the upper right arm are the two best locations that yield the highest information gain (Joshua and Varghese 2013). In this study, the lower arm was not selected for recognition of the activities of interest since it precludes convenient execution of some activities. Consequently, the selection of the upper arm was expected to provide accurate and consistent results compared to other locations on the body. Figure 4 shows some snapshots of the construction workers wearing
mobile phones on their upper arms while performing assigned activities in the experiments conducted in this research.

![Figure 4. Snapshot of construction worker data collection experiments using mobile phones](image)

**Data Collection and Logging**

Technical details of the data collection in terms of the mobile applications, sensors used (accelerometer and gyroscope), and sampling frequency (100 Hz) remained the same as the equipment data collection experiment described in the previous Section. Construction workers were asked to do their assigned activities for a certain period of time while waiting for a few seconds in between each instance of their assigned activities. Each activity was performed by two subjects for later user-independent evaluations. One subject performed only sawing. In this case, the goal of activity recognition was to differentiate between the time they were sawing and the time they were not sawing. Another subject performed hammering and turning a wrench. In this case, the activity recognition was intended to detect the time they were hammering, the time they were turning the wrench, and the time they were not doing any of the two activities. Finally, the last subject was responsible for pushing the wheelbarrow and loading/unloading the sections. Therefore, the activities to be recognized in this case were loading sections into a wheelbarrow, pushing a loaded wheelbarrow, dumping sections from a wheelbarrow, and returning with an empty wheelbarrow. Also, the entire experiment was videotaped for data annotation.

**Data Processing and Classification**

Classifications are conducted in 3 activity categories. Category 1 includes only one distinguishable activity, sawing, to assess the performance of the classifiers in detecting value-adding versus non-value-adding instances in a pool of accelerometer and gyroscope data. The result of classification in this category contributes to the overall performance of the developed activity recognition system when used for productivity measurement. In this category, sawing is categorized against idling. Category 2 includes instances of consecutive hammering and turning a wrench as two adjacent activities with almost similar corresponding movements of the worker’s arm. These two activities are also classified against idling to assess the accuracy of the developed activity recognition system in differentiating between activities that produce similar physical body motions. Finally, in category 3, four activities that produce different distinguishable body movements are categorized. These activities include loading sections into...
a wheelbarrow, pushing a loaded wheelbarrow, dumping sections from a wheelbarrow, and returning an empty wheelbarrow, that were also categorized against idling. Similar to the equipment experiments, 128 data points were segmented in one window and 50% overlap for the adjacent windows is considered. In this experiments, the extracted features are as follow: mean, maximum, minimum, variance, RMS, IQR, and correlation between each two pairs of axes comprised the seven time-domain features and spectral energy and entropy were the two frequency domain features. Considering data collection in three axes of the two sensors and nine independent features extracted per sensor per axis, a total of 54 features were extracted from all collected data.

Results of Workers’ Activity Recognition

The performance of the classifiers is assessed in two ways. First, the training accuracy of each classifier was calculated. This means that all collected data points were used for both training and testing which provided an overall insight into the performance of a host of classification algorithms in recognizing construction worker activities using accelerometer and gyroscope data. Next, a more robust approach in evaluation of classifiers was adopted. In particular, 10-fold stratified cross validation was used and the results of the 10 replications of the training and testing were averaged out to report the overall accuracy.

The classification accuracies are reported for 3 activity categories listed in Table 3. The following activity codes are used in reporting the results: in the first category, activity sawing (SW) and being idle (ID) are classified. In the second category, activities hammering (HM), turning a wrench (TW), and being idle (ID) are classified. Finally, in the third category classification is performed on the activities loading sections into wheelbarrow (LW), pushing a loaded wheelbarrow (PW), dumping sections from wheelbarrow (DS), returning an empty wheelbarrow (RW), and being idle (ID). Table 3 shows the results of training and 10-fold cross validation classification accuracy the subject performing activities of category 1.

Table 3. Category 1 activities classification accuracy

<table>
<thead>
<tr>
<th>Category 1</th>
<th>ANN</th>
<th>Decision Tree</th>
<th>KNN</th>
<th>Logistic Regression</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>100.00</td>
<td>99.36</td>
<td>98.08</td>
<td>98.72</td>
<td>98.19</td>
</tr>
<tr>
<td>10-fold CV</td>
<td>96.77</td>
<td>96.06</td>
<td>95.95</td>
<td>96.05</td>
<td>96.91</td>
</tr>
</tbody>
</table>

According to Table 3, over 99% training accuracy was achieved in category 1 using ANN classifier. This confirms the hypothesis that IMU data pertaining to a single activity performed by different workers contain highly distinguishable patterns. However, training accuracy is not an appropriate measure to assess the ability of using such data for new instances of the same activity. Nevertheless, the stratified 10-fold cross validation results confirm that regardless of the nature of classification algorithm, a single activity can be recognized with over 96% accuracy using all five classifiers.

Since it is very likely that a construction worker performs more than one highly distinguishable activity at a time, activities performed in category 2 are designed such that they produce almost the same physical arm movement. Table 4 shows the training and 10-fold cross validation classification accuracy results of both subjects performing activities of category 2.
Table 4. Category 2 activities classification accuracy

<table>
<thead>
<tr>
<th>Category 2</th>
<th>ANN</th>
<th>Decision Tree</th>
<th>KNN</th>
<th>Logistic Regression</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>98.62</td>
<td>97.07</td>
<td>93.81</td>
<td>88.14</td>
<td>87.28</td>
</tr>
<tr>
<td>10-fold CV</td>
<td>93.19</td>
<td>85.83</td>
<td>87.80</td>
<td>86.42</td>
<td>85.34</td>
</tr>
</tbody>
</table>

Similar to category 1, the training accuracies are high particularly for the ANN classifier and the decision tree. While CART decision trees are not very stable and a small change in the training data can change the result drastically as appears in the outcome of the 10-fold cross validation, ANN presents an average of around 90% accuracy for both subjects. This is while all other classification methods performed almost the same with a slight superiority of KNN relative to the other algorithms. This result is particularly important considering the fact that the two activities in category 2 (i.e. hammering and turning a wrench) produce almost similar physical movements in a worker’s arm.

In the third category, a mixture of different distinguishable activities performed by typical construction workers is included to evaluate the performance of the developed activity recognition system in recognizing them. Table 5 shows the training and 10-fold cross validation classification accuracy results of both subjects performing activities of category 3.

Table 5. Category 3 activities classification accuracy

<table>
<thead>
<tr>
<th>Category 3</th>
<th>ANN</th>
<th>Decision Tree</th>
<th>KNN</th>
<th>Logistic Regression</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>94.80</td>
<td>97.11</td>
<td>95.75</td>
<td>90.37</td>
<td>85.82</td>
</tr>
<tr>
<td>10-fold CV</td>
<td>92.01</td>
<td>87.95</td>
<td>90.75</td>
<td>90.75</td>
<td>84.42</td>
</tr>
</tbody>
</table>

According to Table 5, again decision tree gained a high accuracy in training while as expected; its performance is not the same in 10-fold cross validation evaluation. However, except for the decision tree and SVM, all other classifiers, namely ANN, KNN, and logistic regression resulted in around 90% average accuracy for both subjects. Similar to the other two categories, the feedforward back-propagation implementation of the ANN resulted in the highest accuracy among all.

**SUMMARY AND CONCLUSIONS**

The goal of the research presented in this paper was to investigate the prospect of using built-in smartphone sensors as ubiquitous multi-modal data collection and transmission nodes in order to detect detailed construction equipment activities. In spite of its importance, automated recognition of construction worker and equipment activities on the jobsite has not been given due attention in CEM literature. The discovered process-level knowledge can provide a solid basis for different applications such as productivity improvement, safety management, and fuel use and emission monitoring and control. In addition, this methodology can serve as a basis for activity duration extraction for the purpose of construction simulation input modeling.

In order to study construction equipment activity recognition, a case study of front-end loader was used to describe the methodology for action recognition and evaluate the performance of the developed system. In doing so, several important technical details such as selection of discriminating features to extract different LoDs for classification, and choice of classifier to be
trained were investigated. Results indicated that different equipment actions generate distinct data patterns (i.e. signatures) in accelerometer and gyroscope data. In the 3-class level a high accuracy of as 98.59% can be achieved using ANN. With the same choice of classifier, 81.30% and 86.09% accuracy was achieved for 4- and 5-class levels.

In case of workers’ activity recognition, built-in sensors of ubiquitous smartphones have been employed to assess the potential of wearable systems for activity recognition. Smartphones were affixed to workers’ upper arms using armbands, and accelerometer and gyroscope data were collected from multiple construction workers involved in different types of activities. The high levels of training accuracies achieved by testing several classification algorithms including ANN, decision tree, KNN, logistic regression, and SVM confirmed the hypothesis that different classification algorithms can detect patterns that exist within signals produced by IMUs while different construction tasks are performed. Through 10-fold stratified cross validation, algorithms were trained with 90% of the available data and the trained models were tested on the remaining 10%. In different categories of activities, around and over 90% accuracy was achieved. This promising result indicates that built-in smartphone sensors have high potential to be used as integrated data collection and activity recognition platforms in construction environments.

FUTURE WORK
A potential direction for future work in this research will be to explore whether the results achieved so far can be used for automatically extracting process knowledge such as activity durations and precedence logic for the purpose of ubiquitously updating and maintaining simulation models corresponding to field operations. In addition, another branch of future work rooted in the current research is automated identification of unsafe workers’ postures in physically demanding construction activities. Work-related Musculoskeletal Disorder (WMSD), back, knee, and shoulders injuries are among the most common injuries that can be prevented or reduced by complying with Occupational Safety and Health Administration (OSHA) or the National Institute for Occupational Safety and Health (NIOSH) standards and rules (NIOSH 2015; OSHA 1990).

ACKNOWLEDGMENTS
The authors would like to acknowledge the help and support of Hubbard Construction, for providing access to active construction facilities for equipment data collection experiments. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of Hubbard Construction. Also, the authors are grateful to the construction engineering research laboratory students who assisted in the data collection process.

REFERENCES


Akhavian, Brito, Behzadan


