Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: 1547-2450 (Print) 1547-2442 (Online) Journal homepage: http://www.tandfonline.com/loi/gits20

Detection and Classification of Vehicles by Measurement of Road-Pavement Vibration and by Means of Supervised Machine Learning

Markus Stocker, Paula Silvonen, Mauno Rönkkö & Mikko Kolehmainen

To cite this article: Markus Stocker, Paula Silvonen, Mauno Rönkkö & Mikko Kolehmainen (2016) Detection and Classification of Vehicles by Measurement of Road-Pavement Vibration and by Means of Supervised Machine Learning, Journal of Intelligent Transportation Systems, 20:2, 125-137, DOI: 10.1080/15472450.2015.1004063

To link to this article: https://doi.org/10.1080/15472450.2015.1004063

Accepted author version posted online: 06 Feb 2015.
Published online: 22 May 2015.

Journal of Intelligent Transportation Systems, 20(2):125–137, 2016

ISSN: 1547-2450 print / 1547-2442 online
DOI: 10.1080/15472450.2015.1004063

Taylor & Francis
Taylor & Francis Group

Detection and Classification of Vehicles by Measurement of Road-Pavement Vibration and by Means of Supervised Machine Learning

MARKUS STOCKER,[1] PAULA SILVONEN,[2] MAUNO RÖNKKÖ,[1] and MIKKO KOLEHMAINEN[1]

[1]Department of Environmental Science, University of Eastern Finland, Kuopio, Finland
[2]VTT Technical Research Centre of Finland, Helsinki, Finland

Road vehicle detection and, to a lesser extent, classification have received considerable attention, in particular for the purpose of traffic monitoring by transportation authorities. A multitude of sensors and systems have been developed to assist people in traffic monitoring. Camera-based systems have enjoyed wide adoption over the last decade, partially substituting for more traditional techniques. Methods based on road-pavement vibration are not as common as camera-based systems. However, vibration sensors may be of interest when sensors must be out of sight and insensitive to environmental conditions, such as fog. We present and discuss our work on detection and classification of vehicles by measurement of road-pavement vibration and by means of supervised machine learning. We describe the entire processing chain from sensor data acquisition to vehicle classification and discuss our results for the task of vehicle detection and the task of vehicle classification separately. Using data for a single vibration sensor, our results show a performance ranging between 94% and near 100% for the detection task (1340 samples) and between 43% and 86% for the classification task (experiment specific, between 454 and 1243 samples).

Keywords Digital Signal Processing; Machine Learning; Road Vehicle Detection and Classification; Vibration Sensors

INTRODUCTION

We describe the tasks of (a) detection and (b) classification of road vehicles by measurement of road-pavement vibration and by means of supervised machine learning, in particular multilayer perceptron (MLP) feed-forward artificial neural networks (ANN). The presented experiments are part of a platform for the monitoring of an operational environment. The platform is used to develop information systems for situation awareness and security applications. It is located at the training area of the Emergency Services College, Kuopio, Finland. It includes sensors and systems to detect chemical emissions in water, outdoor and indoor air, as well as sensors intended for vehicle access control.

Address correspondence to Markus Stocker, Department of Environmental Science, University of Eastern Finland, P.O. Box 1627, FI-70211, Kuopio, Finland. E-mail: markus.stocker@uef.fi

There has been considerable research into vehicle detection and/or classification (Gupte, Masoud, Martin, & Papanikolopoulos, 2002), in particular for the purpose of traffic monitoring in intelligent transportation systems (Mimbela & Klein, 2000). A number of techniques and sensors have been proposed, ranging among vision-based detectors (Gupte et al., 2002; Hsieh, Yu, Chen, & Hu, 2006; Lai, Fung, & Yung, 2001; Wu, QiSen, & Mingjun, 2001); inductive loop detectors (Gajda, Sroka, Stencel, Wajda, & Zeglen, 2001; Mimbela & Klein, 2000), acoustic signature analysis (Kozhisseri & Bikdash, 2009; Nooralahiyan, Dougherty, McKeown, & Kirby, 1997; Mimbela & Klein, 2000; Takechi, Sugimoto, Mandono, & Sawada, 2004), vehicle axle sensors (Gajda et al., 2001), pneumatic road tube (Mimbela & Klein, 2000), piezoelectric sensors

(Mimbela & Klein, 2000), millimeterwave (microwave radar) sensors (Mimbela & Klein, 2000; Nüssler, Essen, & Buth, 2005), accelerometer sensors (Bajwa, Rajagopal, Varaiya, & Kavaler, 2011), laser range sensors (Harlow & Peng, 2001), magnetic field sensors (Haoui, Kavaler, & Varaiya, 2008; Mimbela & Klein, 2000), weigh-in-motion systems (Mimbela & Klein, 2000), and infrared and ultrasonic sensors (Mimbela & Klein, 2000). In particular for the purposes of driver assistance (Sun, Bebis, & Miller, 2004, 2006), traffic monitoring (Hsieh et al., 2006), and auto-toll systems (Lai et al., 2001) the application of vision-based vehicle detection has received considerable attention. Vehicle detection and classification have been of importance also to security purposes (Jackowski & Wantoch-Rekowski, 2005; Nüssler et al., 2005).

Each of the techniques has advantages and disadvantages (Mimbela & Klein, 2000). For instance, inductive loop detectors are insensitive to weather conditions (Harlow & Peng, 2001) but are intrusive, have high installation and maintenance costs (Bajwa et al., 2011), and may lead to inaccurate results (Kozhisseri & Bikdash, 2009), for example, in detecting truck axles as separate vehicles (Harlow & Peng, 2001). In contrast, vision-based systems are sensitive to environmental conditions and must be in line of sight (Kozhisseri & Bikdash, 2009) but are nonintrusive (Bajwa et al., 2011).

Being a more difficult task (Nooralahiyan et al., 1997), vehicle classification has, compared to vehicle detection, received less attention (Gupte et al., 2002), even though the class of a vehicle is considered to be an important parameter for road-traffic monitoring or for access control (Gajda et al., 2001).

Using vibration (accelerometer) sensors to monitor pavement acceleration and magnetometer sensors for vehicle detection, Bajwa et al. (2011) propose a novel algorithm to estimate axle count and spacing for trucks. Based on a distribution of estimated axle spacing, the authors found in their data three distinctive clusters that can be attributed to groups of trucks. Hostettler et al. also studied vehicle axles detection using road-pavement vibration (Hostettler, Birk, & Nordenvaad, 2010). The method presented in this article relates to that discussed by Bajwa et al. and Hostettler et al. in that the three studies rely on road-pavement vibration induced by vehicles. Contrary to Bajwa et al. and Hostettler et al., who developed "white box" models, we use supervised machine learning to classify the frequency profile of measured vibration and, thus, adopt "black-box" modeling. Supervised machine learning has been used for road vehicle detection and classification (Jackowski & Wantoch-Rekowski, 2005; Nooralahiyan et al., 1997; Wu et al., 2001).

Jackowski and Wantoch-Rekowski (2005) discuss the problem of using artificial neural networks for military vehicle classification using ground vibration. The authors compute the coefficients of a five-order linear predictive coding (LPC) (Gray, 2010) model from the vibration signal. LPC is a technique used in speech processing (Nooralahiyan et al., 1997). The five coefficients are then used to train a multilayer artificial neural network with two outputs, one for light and one for heavy vehicles (> 12,000 kg).

Wu et al. (2001) extract a vehicle model consisting of 29 structural parameters corresponding to vehicles in images captured by CCD cameras. The authors compute a set of vertices from vehicle structural parameters and the distance between pairs of vertices, for example, the distance between two vertices for the vehicle roof plane width. The computed distances between pairs of vertices are used as features in training a multilayer perceptron (MLP) feed-forward artificial neural network.

Nooralahiyan et al. (1997) use a directional microphone, LPC to extract features, and a time-delay neural network (TDNN) to classify road vehicles based on their (speed-independent) acoustic signature. Aside from LPC, the authors also investigated auditory model processing (a computational model of hearing) and Fourier-transform digital signal processing techniques. In addition to the supervised TDNN, the unsupervised Kohonen network (self-organizing map, Kohonen, 1982) was also tested.

We investigate methods based on vibration data because, contrary to vision-based techniques, vibration sensors can be deployed in an area such that they are out of sight. Further, they are insensitive to environmental conditions that impair visibility, such as fog or darkness. Our experimental setting is one of low-traffic, that is, occasional events, and a restricted-access area. Hence, we argue that the discussed materials and methods are of particular interest to security applications, such as monitoring of activity in, and detection of unauthorized access to, restricted areas. The main contributions of this study are that (a) we show how road-pavement vibration can be utilized to reliably detect road vehicles, and (b) to a lesser extent classify vehicles, and (c) we describe the entire processing chain from sensor data acquisition to validated classification results for (d) a relatively large study including almost 2000 events acquired from 5 billion vibration measurement values and a total amount of 1.7 TB of data. Limitations of this study include (a) the use of only one sensor, (b) method development and evaluation specifically for a low-traffic road, and (c) the use of vibration data only for summer road conditions.

MATERIALS AND METHODS

We used a CEF C3M01 vibration sensor developed by Control Express Finland (CEF) Oy for condition monitoring and machinery maintenance. (CEF C3M01 vibration sensors are now manufactured by Webrosensor Oy as WBS CM301.) The sensor was installed at the side of a road (62.83124357N 27.51529644E, WGS84), mounted on a metal bar that penetrated approximately 1m horizontally into the ground about 0.3m below the paved road surface. Figure 1 pictures how the sensor was installed at the side of the road. Figure 2 shows the camera perspective over the test area; the location of the vibration sensor approximately coincided with the position of the fire

Figure 1 The sensor was mounted on a metal bar that penetrated approximately 1 m horizontally into the ground about 0.3 m below the paved road surface. In this image, the pavement extended from the upper right corner. Visible are the sensor, the left end of the metal bar, the Ethernet cable, and a roadside marker that located the sensor in the environment.

truck on the road. CEF C3M01 vibration sensors measure the acceleration of monitored objects. Specifically, in this study the sensor was used to measure road-pavement vibration, in particular also vibration induced by vehicles. The sensor was connected to (and was powered by) an Ethernet network. It served sampled data encoded as Waveform Audio File Format (WAVE) (IBM Corporation & Microsoft Corporation, 1991) over HTTP (Internet Engineering Task Force, 1999).

To visually monitor the road we employed an AXIS 211W wireless network camera with an outdoor antenna kit AXIS 211W (62.83069255N 27.51415649E, WGS84). The camera served JPEG (ISO/IEC JTC 1/SC 29, 2009) images (640 × 480 pixels) over HTTP. Note that images were acquired for the sole purpose of creating a labeled data set and thus are not used in machine learning.

Figure 2 Example camera image that shows the test area. The vibration sensor was installed at the right side of the road approximately where the moving fire truck is located in the picture. The camera was positioned on top of a viewpoint tower located near the monitored road section.

Software was developed in Java. We used the Modular Audio Recognition Framework (MARF, version 0.3.0) for digital signal processing (bandpass filter and FFT). For machine learning we used WEKA (version 3.6.5) (Hall et al., 2009). We have tested multiple data sampling methods, preprocessing techniques, and classification algorithms. Computations were performed on an Intel Core 2 central processing unit (CPU) at 3.16 GHz with 8 GB RAM and an Ubuntu 11.04 Linux operating system. Thus, sensor data were processed by a single computer and time stamps assigned to data are assumed to be synchronized. Sensor data were stored to a 2-TB hard disk by Samsung Electronics Co., Ltd (model HD204UI).

The methods used to build training data sets from vibration data are presented next. Figure 3 gives a diagram that summarizes the steps.

Sensor Data Acquisition

We developed software to continuously acquire, parse, and store both vibration and camera data. Measurement sampling rate for the CEF C3M01 vibration sensor was set to 2000 Hz and measurement session duration to 600 s. A new connection was initialized automatically at the close of a measurement session. Vibration data were processed by means of a customized WAVE parser. After parsing, vibration data were available as a stream of tuples (t, v) where t is a time stamp and v a double value for the measured acceleration at t. Apache Cassandra with a customized data model for time-series data, keyed by time UUID (Internet Engineering Task Force, 2005) at microsecond resolution, was used to persistently store tuples (t, v) of vibration data. JPEG images served by the AXIS 211W camera at a variable 1–3 frames per second (depending on the available network bandwidth) were stored as individual files to disk. The time upon writing, in milliseconds, was used as the image file name.

Vibration Event Localization

To software-assist the localization of vibration event(s) of interest, E_V, that is, road-pavement vibration induced by vehicles, we developed software that, in essence, implemented an iterative window operation over vibration data stored in Apache Cassandra. The size of the window, W, was set to $2^{14} = 16,384$ measurement values, that is, 8.192s (at 2000 Hz). Figure 4a shows an example of raw vibration data for a window that contains an event-of-interest. In comparison, Figure 4b shows an example of raw vibration data for a window that does not contain an event-of-interest. We iteratively queried Apache Cassandra

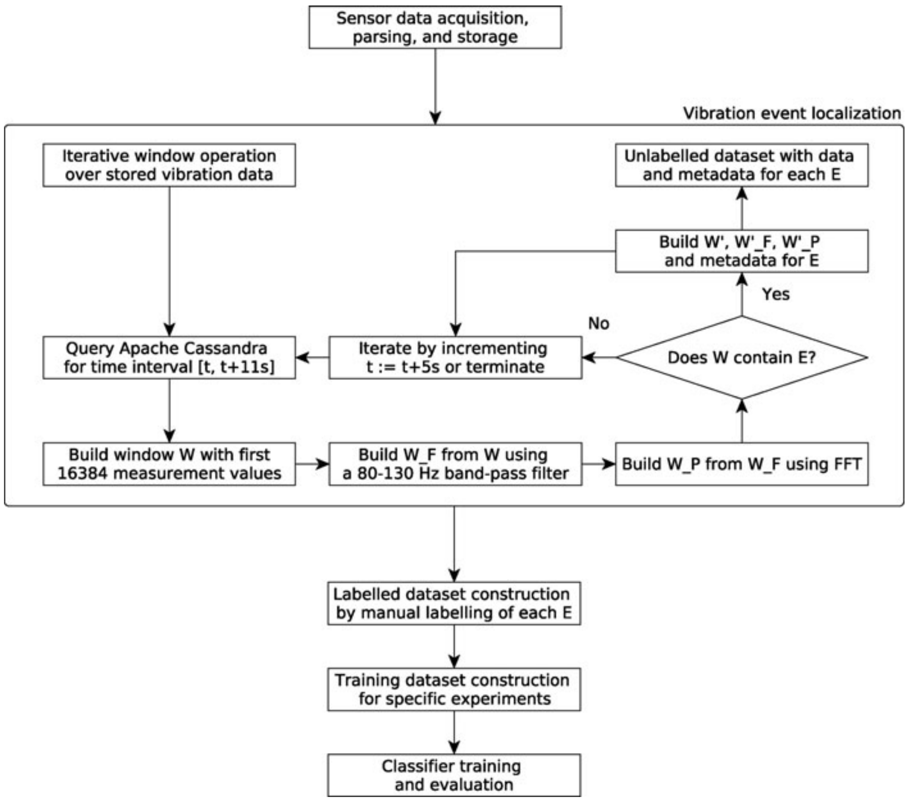

Figure 3 Overview diagram of the presented method we adopted to build training data sets, from vibration data, for classifiers that were trained, in a supervised manner, and evaluated for specific experiments.

for data in the time interval $[t_i, t_i + 11s]$. The first 16,384 measurement values formed W_i. At each iteration, t_i was increased by 5s, that is, $t_{i+1} = t_i + 5s$, in order to form the next query interval $[t_{i+1}, t_{i+1} + 11s]$, of which the first 16,384 measurement values formed W_{i+1}.

For each W we applied a bandpass filter between 100 Hz and 160 Hz. We refer to a filtered window as W_F. In preliminary work, using manual spectrogram analysis of multiple windows W_i containing an event of interest, we discovered (a) strong spectral density at 50 Hz, attributable to electric power, and (b) spectral density for frequencies corresponding to road-pavement vibration between approximately 100 Hz and 160 Hz. Thus, we applied a bandpass filter to enhance the signal between 100 Hz and 160 Hz. Experimentally we have seen that the filtered signal in W_F shows E_V much clearer than in the corresponding window W for the raw vibration data. Figure 5a shows the bandpass filtered window W_F containing an event of interest. In comparison, Figure 5b shows the bandpass filtered window W_F not containing an event of interest. We then computed W_P for W_F, that is, the frequency profile between 100 Hz and 160 Hz, according to the discrete Fourier transform as computed by the MARF implementation of the fast Fourier transform (FFT) algorithm. Figure 6a shows the frequency profile W_P for the window W_F containing an event-of-interest. In comparison, Figure 5b shows the frequency profile W_P for the window W_F not containing an event-of-interest.

Our software decided based on W_P whether or not W *may* contain an event of interest. Specifically, the implementation sorted the values of W_P into ascending numerical order and marked W as possibly containing an event of interest if the last, the fifth-last, and the 10-last values were above the threshold value 2. This threshold was determined experimentally in preliminary work by inspecting numerous profiles with, and profiles without, an event of interest (see Figure 6a and Figure 5b). Therefore, our method was unable to discern between an event of interest, E_V, for vibration induced by vehicles (i.e., true positives) and an event of noise, E_N, for vibration not induced by vehicles (i.e., false positives). Thus, the method only determined (automatically) that W contained an event $E = \{E_V, E_N\}$.

For each W marked as containing an event E, we defined, and queried for, a new window W' such that the variation in W' corresponding to E was approximately centered. We computed the filtered window W'_F and the frequency profile W'_P as already described. We stored the numerical values corresponding to W', W'_F, and W'_P to text files. Finally, we generated metadata for E to keep track of the text file names corresponding to W', W'_F,

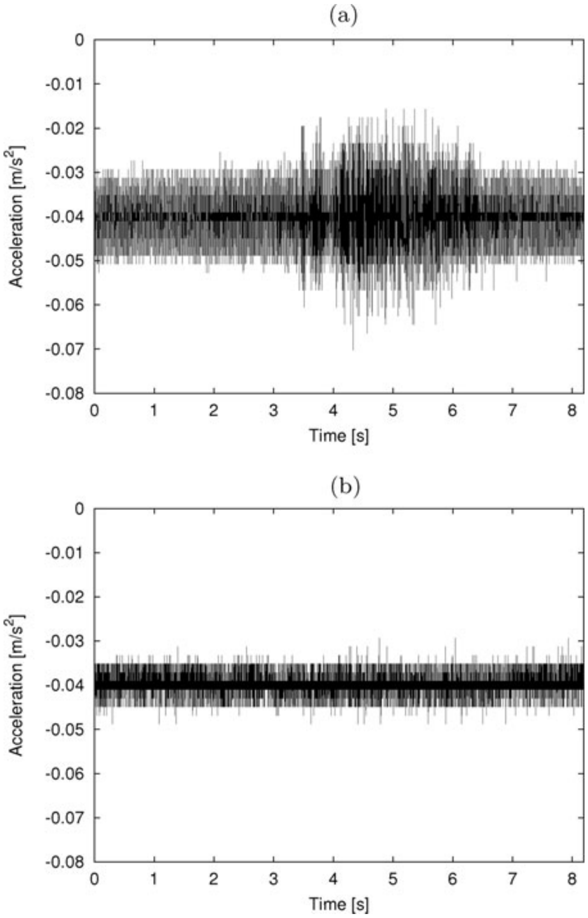

Figure 4 Raw vibration sensor data for a window (not) containing an event of interest. (a) Raw vibration sensor measurement data for a window W consisting of 16,384 measurement values and containing an event-of-interest E_V. (b) Raw vibration sensor measurement data for a window W consisting of 16,384 measurement values and not containing an event of interest E_V.

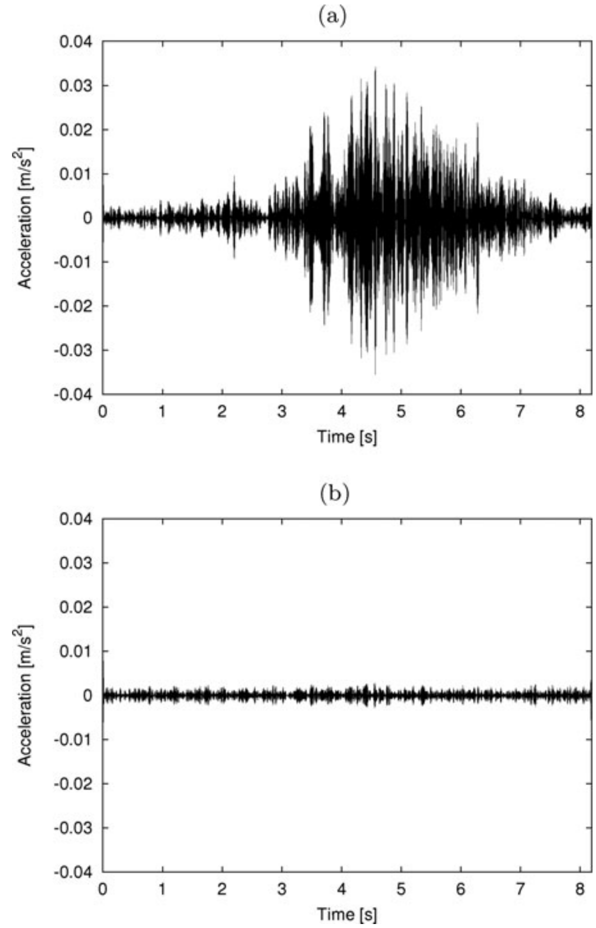

Figure 5 Filtered vibration sensor data for a window (not) containing an event of interest. (a) Window W_F resulting from the application of a MARF bandpass filter between 100 Hz and 160 Hz to W containing an event of interest E_V. (b) Window W_F resulting from the application of a MARF bandpass filter between 100 Hz and 160 Hz to W not containing an event of interest E_V.

and W'_P; time information for the interval $[t_i, t_j]$ defined by W'; time information for E, which is approximately $t_i + 4s$; and the camera image file names that overlap $[t_i, t_j]$. Data and metadata for each E formed an unlabeled data set.

Labeled Data Set Construction

In a manual postprocessing step, the metadata for each E were extended by a label, based on the camera data corresponding to E. Thus, with this step we discerned E_V and E_N. Events E_V were labeled according to the visually identified vehicle, such as `personal-car`, `bucket-digger`, or `fire-truck`. In contrast, events E_N were labeled as `no-event`. In addition, for each E_V we added metadata on road condition (dry/rain); on whether the vehicle came around the corner (yes/no) from/to a side-road (see Figure 1); and on what road side (left/right) the vehicle found itself (relative to the camera) while crossing the sensor. A comprehensive list of vehicle labels is given in Table 1. In addition to data and metadata for W_i containing an event E, we also included data and metadata for W_i not containing an event E (labeled `background`).

Training Data Set Construction

The labeled data set was used to automatically construct experiment-specific training data sets, used to train classifiers and to evaluate their classification performance. Somewhat different approaches were adopted for vehicle detection and vehicle classification. For vehicle detection we selected samples for two training classes, that is, the class `vehicle` and the class `no-vehicle`, such that the two training classes were balanced in number of samples (see Table 2). In contrast, for vehicle classification we selected samples for an experiment-specific number m of training classes, each consisting of a class-specific set of vehicle labels, to form m (balanced) training classes (see Table 4, shown later).

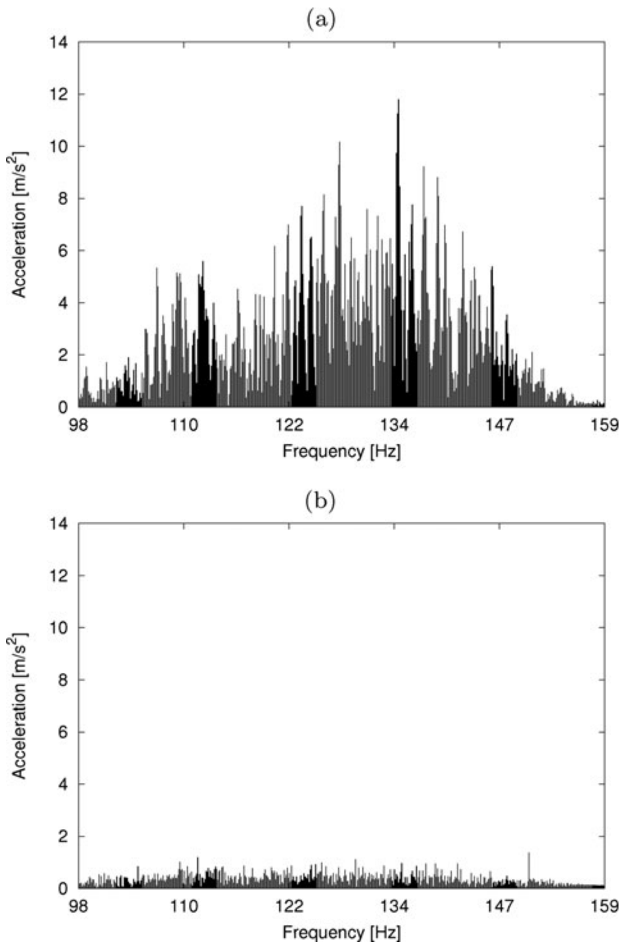

Figure 6 Frequency profile for a window (not) containing an event of interest. (a) Frequency profile W_P with components between 100 Hz and 160 Hz resulting from the application of the discrete Fourier transform to W_F containing an event of interest E_V. (b) Frequency profile W_P with components between 100 Hz and 160 Hz resulting from the application of the discrete Fourier transform to W_F not containing an event of interest E_V.

Table 1 Labeled data set description including unique labels used and corresponding number of samples.

Label	Number	Comment
Motorcycle	8	
Mini-lifter	39	Small-sized vehicle similar to forklift trucks, typically found in warehouses
Mini-cleaner	61	Small-sized cleaning vehicle with single horizontally rotating sweeper in front
Personal-car	285	
Van	167	
Fire-van	81	
Ambulance	30	
Pickup-truck	80	Small-sized vehicle capable of transporting light cargo
Fire-truck	57	
Truck	221	Generic truck capable of transporting cargo
Bucket-digger	103	
Teboil-truck	8	Trucks of the Finnish oil company Teboil
Long-vehicle	103	Trucks with a trailer of size as to give the vehicle approximately twice the load capacity
Background	457	No vehicle (background vibration)
No-event	196	Unexplained, E_N events
Worker	14	Walking people, E_N events
No-images	1	Missing camera images
Total	1911	

Classifier Training and Evaluation

Experiment-specific training data sets consisted of data rows for each selected sample and were compliant with the WEKA ARFF file format. Each row consisted of 500 attributes for the data of W'_P and 1 attribute for the corresponding, experiment-specific, training class (e.g., `no-vehicle`, `van-like`, `heavy`). Training data sets were loaded in WEKA, and principle component analysis (PCA) was used in preprocessing to compute the scores corresponding to the first 10 principal components. (This threshold was determined experimentally as having good classification performance improvement.) We thus reduced the dimensionality from 500 to 10. For both vehicle detection and vehicle classification tasks, the artificial neural network consisted of 10 input neurons and 1 hidden layer with $\frac{I_N+O_N}{2}$ hidden neurons, where $I_N = 10$ is the number of input neurons and O_N is the number of output neurons, that is, the number of experiment-specific training classes. The learning rate and momentum for the back-propagation algorithm were set to 0.3 and 0.2, respectively. The number of epochs was set to 500. (Note that these settings corresponded to default WEKA configuration, which resulted in being fairly reasonable with respect to classification performance.) We used 10-fold cross-validation, meaning that the training data set was randomly partitioned into 10 disjoint and equal-sized folds, and for each fold a classifier was trained using the other 9 folds and then tested on the fold. Thus, samples were of the same data set but at each iteration in cross-validation the samples used in training were different

Table 2 Vehicle detection task training dataset description including training classes and corresponding number of samples.

Label	Training class	Number
Motorcycle	Vehicle	687
Mini-lifter		
Mini-cleaner		
Personal-car		
Van		
Fire-van		
Ambulance		
Pickup-truck		
Fire-truck		
Truck		
Bucket-digger		
Teboil-truck		
Long-vehicle		
Background	No-vehicle	653
No-event		
Total		1340

Table 3 Detailed accuracy by class and confusion matrix for the vehicle detection task.

(a) Detailed accuracy by class.

TP rate	FP rate	Precision	Recall	Training class
0.934	0.054	0.943	0.934	No-vehicle
0.946	0.066	0.938	0.946	Vehicle
0.94	0.06	0.94	0.94	Weighted average

(b) Confusion matrix.

a	b	Training class (Number)
610	43	a = No-vehicle (653)
37	**650**	b = Vehicle (687)

from the samples used in testing. The 10 intermediate results for the folds were averaged to produce the results discussed in the next section.

RESULTS

We have collected and stored the data from both the CEF C3M01 vibration sensor and the AXIS 211W camera for approximately 710 h of measurement. After each measurement session (typically a working day) we recorded the total amount of data stored for both sensors as well as the number of measurements stored for the vibration sensor. The total amount of stored data was approximately 1.7 TB. The number of vibration measurement values stored was approximately 5 billion. Collecting vibration data at 2000Hz for 710 h theoretically leads to 5,112,000,000 measurement values. Experimentally, we recorded a total of 5,004,130,000. Thus, in our experiments we have a gap of 107,870,000 measurements, which corresponds to approximately 15 h. This gap is due to a combination of factors including approximate calculations, sensor reconnection every 10 min, and possibly uncovered sensor downtime. The inconsistency in numbers of collected and theoretical measurements has practical consequences in that (a) we may identify vehicles in camera images but, as we lack of the corresponding vibration data, we must discard valuable training samples; (b) on the other hand, in the case of missing images we may discover an event in vibration data but are unable to label the event accordingly, leading to further loss in training samples; (c) in a real-time context, (temporary) unavailability of vibration data can mean that we miss out on events; and (d) gaps in vibration data of a few seconds can influence window processing, as we may have insufficient data or we may be processing data that are not sampled at a constant interval.

Our labeled data set contains 1911 samples. Table 1 provides an overview. The data set consists of one sample with no images (`no-images`); 210 samples with events of type E_N, that is, events for vibration not induced by vehicles but rather either by unknown factors (`no-event`, 196) or by people passing by the sensor on foot (`worker`, 14); and 457 samples not containing an event E (`background`). Thus, the data set consists of 1243 samples that contain an event E_V.

The training data set used for the vehicle detection task contains 1340 samples and consists of two classes. Table 2 provides an overview. The first class, `vehicle`, consists of 687 samples that contain an event E_V, that is, a detected vehicle. The second class, `no-vehicle`, consists of 653 samples that do not contain an event. Note that we limited the number of samples per training class to 687 in order to keep the classes approximately balanced. Cross-validation resulted in 1260 correctly classified instances (94%). Table 3 shows the detailed accuracy by class and confusion matrix for the detection task as reported by WEKA.

As we noted in the second section, the training data set used for the vehicle classification task is specific to individual experiments. We performed three distinct experiments, named 2-class, 4-class, and 13-class. For the 2-class experiment we map the 13 vehicle labels to 2 training classes, one for `light` and the other for `heavy` vehicles, by weight. This experiment is balanced, meaning that the two training classes consist of approximately the same number of samples (limited to 500 samples per training class) with a total of 992 samples. For the 4-class experiment we map a subset of the 13 vehicle labels to 4 training classes, that is, `van-like`, `truck-like`, `bucket-digger`, and `long-vehicle`. This experiment is also balanced (limited to 120 samples per training class) with a total of 454 samples. Finally, for the 13-class experiment we map the 13 vehicle labels to (corresponding) 13 training classes. This last experiment is unbalanced, meaning that the number of samples per training class is highly variable (minimum 8; maximum 285) with a total of 1243 samples. We refer to Table 4 for detailed information on the vehicle labels each training class consists of and the corresponding number of samples, for each of the three experiments. We found the classification performance as per cross-validation to be 86% for the 2-class experiment, 64% for the 4-class experiment, and 43% for the 13-class experiment. Table 5 shows the detailed accuracy by class and Table 6 the confusion matrix for the three experiments, respectively, as reported by WEKA.

Over the past decades, machine-learning and data-mining communities have developed a large number of methods for unsupervised and supervised machine learning (Mitchell, 1997). In addition to MLP, we investigated the classification performance for several other machine learning algorithms, including `RBFNetwork`, `NaiveBayes`, and `SVM`. Table 7 provides an overview. We found that, everything else being equal, the classification performance of different machine learning algorithms is roughly comparable. For instance, for the vehicle detection task the classification performance was for `NaiveBayes` 75%, `RBFNetwork` 86%, and `SVM` 93% (`MLP` 94%), while for the 4-class vehicle classification task the classification performance was for `NaiveBayes` 55%, `RBFNetwork` 60%, and `SVM` 65% (`MLP` 64%).

We also investigated the effect of data preprocessing on classification performance, as well as the characteristics of training datasets. PCA generally improves classification performance

Table 4 Vehicle classification task dataset description including training class and corresponding number of samples for the 2-class, 4-class, and 13-class experiments.

(a) 2-Class experiment (balanced).			(b) 4-Class experiment (balanced).			(c) 13-Class experiment (unbalanced).		
Label	Training class	Number	Label	Training class	Number	Label	Training class	Number
Motorcycle	Light	500	Motorcycle			Motorcycle	Motorcycle	8
Mini-lifter			Mini-lifter			Mini-lifter	Mini-lifter	39
Mini-cleaner			Mini-cleaner			Mini-cleaner	Mini-cleaner	61
Personal-car			Personal-car			Personal-car	Personal-car	285
Van			Van	Van-like	120	Van	Van	167
Fire-van			Fire-van			Fire-van	Fire-van	81
Ambulance			Ambulance			Ambulance	Ambulance	30
Pickup-truck			Pickup-truck			Pickup-truck	Pickup-truck	80
Fire-truck	Heavy	492	Fire-truck	Truck-like	120	Fire-truck	Fire-truck	57
Truck			Truck			Truck	Truck	221
Bucket-digger			Bucket-digger	Bucket-digger	103	Bucket-digger	Bucket-digger	103
Teboil-truck			Teboil-truck	Long-vehicle	111	Teboil-truck	Teboil-truck	8
Long-vehicle			Long-vehicle			Long-vehicle	Long-vehicle	103
Total		992	Total		454	Total		1243

by approximately 10%. Not surprisingly, the characteristics of the training data set have a considerable effect on classification performance. Obviously, one factor is the number of training classes. As can be seen in Table 5, the classification performance is notably different between the 2-class (86%), the 4-class (64%), and the 13-class (43%) vehicle classification task experiments. Balancing the data set such that all training classes have approximately the same number of samples also affects the classification performance. In our experiments we used balanced training data sets, except for the 13-class vehicle classification task experiment. To demonstrate the effect of balancing, consider for instance an unbalanced data set for the 4-class vehicle classification task experiment, consisting of 770 samples (having `van-like` and `truck-like` almost three times more often represented than `long-vehicle` and `bucket-digger`). The classification performance is then 72% (compared to 64% using a balanced data set). Indeed, an algorithm that always predicts (one of) the most represented class(es) may perform very well on an evaluation data set in which one class (or a few classes) is represented in much higher numbers than the remaining classes. Hence, in our experiments we have kept the training and evaluation data sets balanced (where possible).

DISCUSSION

We have described the entire processing chain from sensor data acquisition to vehicle classification for the purposes of detection and classification of vehicles by measurement of road-pavement vibration and by means of supervised machine learning, in particular MLP. We have presented and discussed our results for classification performance and compared them with the results of similar studies published in the literature.

The platform monitoring infrastructure consists of, among other sensors, a CEF C3M01 vibration sensor installed in the ground on the side of a paved road, and the necessary communication network links. In addition, a camera is employed to enable visual monitoring of vehicles. We have discussed the methods and software used to persistently store high-frequency and high-volume sensor data. Such historical data are used for vibration data analysis, as well as to generate data and metadata on events of interest required to build training data sets for supervised machine learning. We have presented our results for both vehicle detection and vehicle classification tasks. Classification performance was evaluated and presented for three experiments with experiment-specific mappings of vehicle labels to training classes.

The performance of the binary vehicle detection task is, at 94%, good. Considering the visibly different frequency profiles of windows that contain an event (Figure 6a) and windows that do not (Figure 5b) this may not be surprising. Indeed, we might ask why the classifier fails in 6% of the cases. As summarized in Table 2, the `no-vehicle` training class consists of both `background` and `no-event` labels. The label `no-event` consists of samples that were (automatically) marked as containing an event E (possibly E_V). However, in manual postprocessing, no vehicle was visible in the set of images corresponding to E. The cause for the vibration remains unknown, but in the field any object with mechanical interference with the sensor, network cable, or metal bar is likely to induce vibration. We performed a second test for the vehicle detection task, and this time we excluded samples labeled `no-event`. In this setting, the training class `vehicle` consisted of 460 samples and the training class `no-vehicle` consisted of 457 samples. Note that we limited the number of samples per class to 460 in order to keep the training classes balanced. Cross-validation resulted in 915 correctly classified instances (99.8%). Two `vehicle` samples were wrongly classified as `no-vehicle`. We performed the same test also using a 66% percentage split for validation, in which case 100% of the validation samples were correctly classified.

Table 5 Detailed accuracy by class for the three vehicle classification task experiments.

(a) 2-Class experiment.

TP rate	FP rate	Precision	Recall	Training class
0.854	0.138	0.859	0.854	Heavy
0.862	0.146	0.857	0.862	Light
0.858	0.142	0.858	0.858	Weighted average

(b) 4-Class experiment.

TP rate	FP rate	Precision	Recall	Training class
0.532	0.096	0.641	0.532	Long-vehicle
0.689	0.06	0.772	0.689	Bucket-digger
0.742	0.096	0.736	0.742	Van-like
0.583	0.237	0.47	0.583	Truck-like
0.637	0.125	0.65	0.637	Weighted average

(c) 13-Class experiment.

TP rate	FP rate	Precision	Recall	Training class
0.329	0.113	0.311	0.329	Van
0.067	0.001	0.667	0.067	Ambulance
0.647	0.154	0.477	0.647	Truck
0.705	0.269	0.438	0.705	Personal-car
0	0.001	0	0	Mini-lifter
0	0	0	0	Teboil-truck
0.709	0.047	0.575	0.709	Bucket-digger
0.099	0.028	0.2	0.099	Fire-van
0.437	0.042	0.484	0.437	Long-vehicle
0.131	0.01	0.4	0.131	Mini-cleaner
0	0.001	0	0	Motorcycle
0	0.001	0	0	Pickup-truck
0.088	0.013	0.238	0.088	Fire-truck
0.434	0.115	0.374	0.434	Weighted average

Note. In the 4-class experiment we underscore the relatively low performance of the training classes `long-vehicle` and `truck-like`. We can observe in Figure 6(b) that the `truck-like` and `long-vehicle` classes were often confused. In the 13-class experiment we underscore that the classes that performed relatively well are those that happened to be well represented as training examples in their respective categories. For instance, there were 57 `fire-truck` versus 221 `truck` samples. Incidentally, the `fire-truck` class was most often confused with the `truck` class. The classes `mini-lifter`, `teboil-truck`, and `motorcycle` were not well represented. Finally, `pickup-truck` was practically always confused as `personal-car` or, to a lesser extent, as `van`, both of which were represented considerably better than `pickup-truck` in the training data set.

As shown in Table 4, we tested a number of different vehicle labels to training class mappings: the 2-class, 4-class, and 13-class experiments. The performance of the 4-class vehicle classification task experiment is, at 64%, satisfactory. The 4-class experiment may be a reasonable trade-off between classification performance, number of training classes, and samples per training class. Light vehicles, such as personal cars, are difficult to separate from van-like vehicles. Because the labeled data set consists of many samples for personal car and van-like vehicles, we included only van-like vehicles in the `van-like` training class. Distinguishing between vans, fire vans, and ambulance is not practical using the presented methods. The same is true for fire trucks and other trucks. As can be seen in Table 6, `long-vehicle` and `truck-like` are often confused. In fact, everything else being equal, discarding the `long-vehicle` class from the training data set leads to 78% correctly classified instances using cross-validation. Overall, the practical significance of the 4-class experiment is arguably questionable. First, it requires us to disregard some of the vehicle types. Second, it does not consider personal cars, which is the most frequent vehicle type. However, together with the 2-class and 13-class experiments, the 4-class experiment may support the drawing of a possible performance trend for progressively difficult classification tasks.

In the 13-class experiment some of the training classes are not well represented, such as `motorcycle`. The performance for some of the vehicle classes is rather good, for example, for `truck` and `personal-car`. However, these classes happen to be the top represented training classes in the data set. The class `van` is mostly confused with `personal-car`, but `personal-car` not much with `van` (probably due to the better representation of `personal-car`). Some classes are strongly confused, such as `pickup-truck` with `personal-car` or `van` (both of which are better represented than `pickup-truck`).

Jackowski and Wantoch-Rekowski (2005) report a study on the classification of wheeled military vehicles using artificial neural networks and ground vibration data. Given the authors' artificial neural network model with two outputs for light and heavy vehicles, their results are comparable to the 2-class vehicle classification task experiment discussed in this article. In their evaluation, test samples are all correctly classified. However, the validation set consisted of only 14 samples. In our 2-class experiment 14% of test samples were misclassified, that is, on average 1.4 out of 10. Hence, the classification performance reported by Jackowski and Wantoch-Rekowski may be comparable to our results. However, it is important to note that Jackowski and Wantoch-Rekowski conducted the study on wheeled military vehicles, and thus the terms *light* and *heavy* are probably interpreted differently.

Wu et al. (2001) also use artificial neural networks for the purpose of vehicle classification, though by means of vision-based techniques and feature extraction from CCD camera images. The data and dataprocessing methods are, hence, different from the methods presented in this article. Wu et al. evaluated their methods using 300 test samples and reported a 91% classification accuracy on six vehicle classes, ranging from car to large truck. This is considerably better than our results for the 4-class experiment (64%), which may signal that feature extraction from camera images could lead to training data for which it is possible to build artificial neural network models with better generalization.

Nooralahiyan et al. (1997) use artificial neural networks on acoustic signature data for vehicle classification. The authors report 82.4% classification accuracy for test samples and four vehicle classes, ranging from motorcycle to van. This result is

Table 6 Confusion matrix for the three vehicle classification task experiments.

(a) 2-Class experiment.

a	b	Training class (Number)
420	72	a = Heavy (492)
69	**431**	b = Light (500)

(b) 4-Class experiment.

a	b	c	d	Training class (Number)
59	6	0	46	a = Long-vehicle (111)
6	**71**	15	11	b = Bucket-digger (103)
1	8	**89**	22	c = Van-like (120)
26	7	17	**70**	d = Truck-like (120)

(c) 13-Class experiment; a = Van (167); b = Ambulance (30); c = Truck (221); d = Personal-car (285); e = Mini-lifter (39); f = Teboil-truck (8); g = Bucket-digger (103); h = Fire-van (81); i = Long-vehicle (103); j = Mini-cleaner (61); k = Motorcycle (8); l = Pickup-truck (80); m = Fire-truck (57).

a	b	c	d	e	f	g	h	i	j	k	l	m
55	0	14	91	0	0	2	5	0	0	0	0	0
15	**2**	5	5	0	0	0	3	0	0	0	0	0
13	0	**143**	20	0	0	9	2	27	1	1	0	5
39	0	22	**201**	0	0	6	8	5	3	0	1	0
4	0	7	12	**0**	0	5	3	4	1	0	0	3
0	0	4	1	0	**0**	1	0	1	1	0	0	0
3	0	7	12	1	0	**73**	3	1	2	0	0	1
26	1	6	31	0	0	4	**8**	1	0	0	0	4
1	0	48	1	0	0	6	0	**45**	0	0	0	2
5	0	9	24	0	0	11	1	3	**8**	0	0	0
1	0	1	2	0	0	1	1	0	2	**0**	0	0
14	0	4	57	0	0	1	2	0	1	0	**0**	1
1	0	30	2	0	0	8	4	6	1	0	0	**5**

Table 7 Summary of the classification performance results for different machine learning algorithms. The table also shows results of related and comparable studies.

		Classification		
Algorithm	Detection	2-Class	4-Class	6-Class
NaiveBayes	75%		55%	
RBFNetwork	86%		60%	
SVM	93%		65%	
MLP	94%	86%	64%	
Jackowski et al.		100%		
Wu et al.				91%
Nooralahiyan et al.			82%	

also better than our results for the 4-class experiment (64%), which may again signal that feature extraction from acoustic data on vehicles may lead to training data for which artificial neural network models generalize better. However, note that the vehicle classes are not equal and acoustic methods may be particularly suitable for the classification of certain vehicles, such as motorcycles.

Building training data sets for supervised machine learning algorithms such as described here is costly. Unable to provide an accurate account, we simply note that, beyond hardware, software infrastructure must be developed to collect and store high-volume and high-frequency sensor data. Algorithms must be implemented to process data such that information can be extracted from raw sensor data. The number of samples required to reach an acceptable classification performance may lie in the few hundreds, which may not be beyond practicability. However, the requirement of balanced training data sets can pose additional demands on resources. More importantly, however, due to different hardware characteristics, each sensor of a network may need its own training data set, in particular if data cannot be normalized such that they are representative for multiple sensors (of the same type). Further, different environmental conditions may also require specific training data sets. For instance, our data set was created by measurement of road-pavement vibration during summer. During winter, roads in eastern Finland are generally covered with ice and snow. Such changed environmental conditions are likely to affect the signal measured and, hence, the characteristics of the data and training data set generated. Thus, a data set generated by measurement in summer may not be suitable for the training of an artificial neural network aimed at classifying vehicles during winter.

The monitoring presented here is for a setting of low traffic. Further, the labeled data set was built such that windows contain one event. Hence, we cannot assume that the discussed methods perform similarly for busy bidirectional traffic monitoring. This is a clear limitation of the presented approach. However, vibration sensors are insensitive to visibility conditions and they can be deployed in the environment such that they are out of sight. Thus, the methods discussed in this article may be of interest to security applications, specifically the monitoring of activity in, and detection of unauthorized access to, restricted areas. Also, vibration sensors are less intrusive compared to other infrastructure, such as inductive loop detectors, and can thus be deployed more quickly and with lower costs. Perhaps the main advantage of using supervised machine learning algorithms for vehicle detection and classification tasks is that we are not required to specifically extract vehicle features from data, such as the number and distance of axles. This reduces the burden on data processing and analysis, as well as the development of specific software. However, the classification performance of supervised machine learning algorithms is only performing in the field as in validation to the extent that the training set is representative for field conditions. This may be problematic in situations where a system must be quickly deployed into new environments or where the conditions of an environment are frequently changing.

The settings of our study, that is, the type of sensors and the physical installation of sensors in the environment, are (experiment) constraints we were required to work with. The specific installation of sensors most probably has an important effect on the signal strength of road-pavement vibration. As described earlier, our installation was not very intrusive. In particular, road pavement was not disrupted by sensor installation. While more intrusive, an installation whereby vibration sensors are embedded into the road pavement could lead to much stronger signals. The discussed methods could hence lead to better classification performance. We argue that the deployment of sensors in the environment can have a significant impact on the performance of data-driven methods.

Finally, we have used only one vibration sensor in our experiments. With two or more vibration sensors installed at constant relative distance it is possible to approximately compute the velocity, speed, and driving direction of vehicles. Speed is likely to affect measured vibration. Knowing the speed of detected vehicles, vehicle classification may perform differently using classifiers that are trained for specific speed ranges, for example [20–40], [40–60], [60–80] km h^{-1}. Knowing the driving direction of detected vehicles, classifiers could also be trained separately for driving side. In fact, the distance between the vehicle and the sensor is also likely to affect measured vibration.

CONCLUSION

We have described the entire processing chain from sensor data acquisition to vehicle classification for the purposes of detection and classification of vehicles by measurement of road-pavement vibration and by means of supervised machine learning, in particular MLP. We have presented and discussed our results for classification performance and compared them with the results of similar studies published in the literature.

The processing chain consists of numerous steps, including sensor data acquisition, its parsing and storing, the identification of information in raw sensor measurement data, and the semiautomatic construction of a labeled data set to be used for creating training data sets that meet the requirements of specific experiments. We argue that building the required software infrastructure necessary to acquire enough representative data to train classifiers in supervised machine learning, and building a corresponding labeled data set, is costly.

Using data for a single vibration sensor, our results show a classification performance ranging between 94% and near 100% for the vehicle detection task, and between 43% and 86% for the vehicle classification task. As we discussed, the resulting performance is a function of a number of experiment settings, including training data sets, preprocessing methods, number of training classes and their characteristics, and learning algorithms. Furthermore, systems based on supervised machine learning only perform in the field as in validation to the extent that the training set is representative for field conditions. Hence, systems based on supervised machine learning may not scale to large sensor networks that operate in environments with frequently changing conditions.

The work presented in this article is part of a wider system located at the training area of the Emergency Services College, Kuopio, Finland. The platform consists of a number of sensors, including several road-pavement vibration sensors and other sensors, such as acoustic sensors. We envision building on the work presented in this article and employing data of multiple road-pavement vibration sensors as well as complement vibration sensor data with, for example, acoustic sensor data for the purpose of vehicle detection and classification. Such a development may lead to better classification performance and may be a relevant study on sensor data fusion for situation awareness. Placing the camera so that vehicles are better visible could also allow the classification of loaded, half-loaded, and empty trucks.

ACKNOWLEDGMENTS

We thank Jarmo Kastinen at Miktech Mikkeli Technology Centre Ltd., formerly at Lappeenranta University of Technology, and Marko Kohtala at Control Express Oy for providing technical assistance regarding sensors, network infrastructure, and platform software.

FUNDING

This work was supported by Tekes, the Finnish Funding Agency for Technology and Innovation (funding decision number 40075/09).

REFERENCES

Bajwa, R., Rajagopal, R., Varaiya, P., & Kavaler, R. (2011). In-pavement wireless sensor network for vehicle classification. In *2011 10th International Conference on Information Processing in Sensor Networks (IPSN)* (pp. 85–96). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5779067&tag=1

Gajda, J., Sroka, R., Stencel, M., Wajda, A., & Zeglen, T. (2001). A vehicle classification based on inductive loop detectors. In *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, 2001, IMTC 2001* (Vol. **1**, pp. 460–464).

Gray, R. M. (2010). *Linear predictive coding and the Internet Protocol. A survey of LPC and a history of realtime digital speech on packet networks*. Retrieved from http://ee.stanford.edu/∼gray/lpcip.pdf.

Gupte, S., Masoud, O., Martin, R., & Papanikolopoulos, N. (2002). Detection and classification of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, **3**(1), 37–47.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. In *SIGKDD Explorations* (Vol. **11**). New York, NY: ACM.

Haoui, A., Kavaler, R., & Varaiya, P. (2008). Wireless magnetic sensors for traffic surveillance. *Transportation Research Part C: Emerging Technologies*, **16**(3), 294–306.

Harlow, C., & Peng, S. (2001). Automatic vehicle classification system with range sensors. *Transportation Research Part C: Emerging Technologies*, **9**(4), 231–247.

Hostettler, R., Birk, W., & Nordenvaad, M. (2010). Feasibility of road vibrations-based vehicle property sensing. *Intelligent Transport Systems, IET*, **4**(4), 356–364.

Hsieh, J. W., Yu, S. H., Chen, Y. S., & Hu, W. F. (2006). Automatic traffic surveillance system for vehicle tracking and classification. *IEEE Transactions on Intelligent Transportation Systems*, **7**(2), 175–187.

IBM Corporation & Microsoft Corporation (1991). *Multimedia programming interface and data specifications 1.0*. Retrieved from http://www-mmsp.ece.mcgill.ca/documents/AudioFormats/WAVE/Docs/riffmci.pdf

Internet Engineering Task Force. (1999). *Hypertext Transfer Protocol – HTTP/1.1 (RFC 2616)*. Retrieved from https://www.ietf.org/rfc/rfc2616.txt

Internet Engineering Task Force. (2005). *A Universally Unique IDentifier (UUID) URN Namespace (RFC 4122)*. Retrieved from http://www.ietf.org/rfc/rfc4122.txt

Jackowski, J., & Wantoch-Rekowski, R. (2005). Classification of wheeled military vehicles using neural networks. In *18th International Conference on Systems Engineering, 2005, ICSENG 2005* (pp. 212–217). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1562854&tag=1

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59–69.

Kozhisseri, S., & Bikdash, M. (2009). Spectral features for the classification of civilian vehicles using acoustic sensors. In *IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, 2009, CIVVS '09* (pp. 93–100). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4938729

Lai, A., Fung, G., & Yung, N. (2001). Vehicle type classification from visual-based dimension estimation. In *Proceedings, Intelligent-Transportation Systems, 2001, 2001 IEEE* (pp. 201–206). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=948656

Mimbela, L. E. Y., & Klein, L. A. (2000). *A summary of vehicle detection and surveillance technologies used in intelligent transportation systems (tech. rep.)*. Berkeley, CA: UC Berkeley Transportation Library.

Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Nooralahiyan, A. Y., Dougherty, M., McKeown, D., & Kirby, H. R. (1997). A field trial of acoustic signature analysis for vehicle classification. *Transportation Research Part C: Emerging Technologies*, **5**(3-4), 165–177.

Nüssler, D., Essen, H., & Buth, D. (2005). Vehicle classification by vibration analysis using millimeterwave sensors. In *The Joint 30th International Conference on Infrared and Millimeter Waves and 13th International Conference on Terahertz Electronics, 2005, IRMMW-THZ 2005* (Vol. **2**, pp. 349–350). Retrieved from http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1572556

Sun, Z., Bebis, G., & Miller, R. (2004). On-road vehicle detection using optical sensors: A review. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, 2004* (pp. 585–590). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1398966

Sun, Z., Bebis, G., & Miller, R. (2006). On-road vehicle detection: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(5), 694–711.

Takechi, T., Sugimoto, K., Mandono, T., & Sawada, H. (2004). Automobile identification based on the measurement of car sounds. In *Industrial Electronics Society, 2004, IECON 2004, 30th Annual Conference of IEEE* (Vol. **2**, pp. 1784–1789). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1431853

Wu, W., QiSen, Z., & Mingjun, W. (2001). A method of vehicle classification using models and neural networks. In *VTC 2001 Spring. IEEE VTS 53rd Vehicular Technology Conference, 2001* (Vol. **4**, pp. 3022–3026). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=944158