

Research Article

A New Terrain Classification Framework Using Proprioceptive Sensors for Mobile Robots

Kai Zhao, Mingming Dong, and Liang Gu

School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Kai Zhao; zhaokai@bit.edu.cn

Received 25 May 2017; Revised 19 July 2017; Accepted 26 July 2017; Published 6 September 2017

Academic Editor: Oscar Reinoso

Copyright © 2017 Kai Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile robots that operate in real-world environments interact with the surroundings to generate complex acoustics and vibration signals, which carry rich information about the terrain. This paper presents a new terrain classification framework that utilizes both acoustics and vibration signals resulting from the robot-terrain interaction. As an alternative to handcrafted domain-specific feature extraction, a two-stage feature selection method combining ReliefF and mRMR algorithms was developed to select optimal feature subsets that carry more discriminative information. As different data sources can provide complementary information, a multi-classifier combination method was proposed by considering a priori knowledge and fusing predictions from five data sources: one acoustic data source and four vibration data sources. In this study, four conceptually different classifiers were employed to perform the classification, each with a different number of optimal features. Signals were collected using a tracked robot moving at three different speeds on six different terrains. The new framework successfully improved classification performance of different classifiers using the newly developed optimal feature subsets. The greater improvement was observed for robot traversing at lower speeds.

1. Introduction

Mobile robots are increasingly deployed in real-world environments, such as forestry, mining, rescue, site inspection, and space exploration [1]. Due to the huge variety of different scenarios that can be encountered in each operating environment, mobile robots must be prepared to deal with complex, unknown, and even dangerous terrains. While they can traverse hard flat ground safely at high speed, they may experience slippage, sinking, and embedding events in the face of loose slippery terrains [2, 3]. Other surfaces can be bumpy and rocky, which may result in damage to the robot [4]. Consequently, the terrain itself can become a hazard, referred to as a nongeometric hazard [5]. To achieve efficient and safe navigation, a mobile robot should adapt its driving style, control strategy, or path planning strategy to accommodate characteristics of the terrain.

A lot of efforts have been made to explore the applications of terrain perception for mobile robots. At present, the most commonly used sensing modalities are cameras and LADARs. Vision-based methods present a powerful tool to perceive the surrounding environment, by which texture or

color information is utilized to define the terrain. However, it is unreliable as changes in appearance may be caused by factors such as illumination, weather, and camouflaging by leaves [6]. The LADAR-based method emphasizes segmenting terrain from obstacles, not the terrain itself, and it can be affected by factors such as rain, snow, and smog [7]. There are considerable limitations in traditional terrain perception methods and the problem of how to reliably sense terrain properties is a challenging task.

It is well known that human beings can capture information about terrain during walking by sensing it with their feet and by the sound of their footsteps, of particular importance in dark environments [8]. Likewise, acoustic and vibration responses resulting from robot-terrain interactions can be exploited for terrain characterization. In previous studies, vibration-based methods have performed well in classifying ground surfaces with different degrees of coarseness [9, 10]. However, acoustic-based methods are believed to be more suitable in distinguishing material types of different terrains [11]. Moreover, the acoustic-based method is highly sensitive to running water and can be used to alert a robot before driving into a stream. Tactile force/torque (F/T) sensors have

been adopted for legged robots to account for the specific leg-terrain interaction [12].

Iagnemma et al. first proposed the vibration-based method for terrain classification [13]. The basic idea is to use vibrations induced in the wheels, the axes, or the body of the robot to classify the underlying terrain being traversed. The method was performed with a planetary rover, in which the PSD function was used to form the feature vector and PCA was used for dimensionality reduction [14]. Weiss et al. used vibration response collected by a cart with hard wheels to classify the interactions [15]. Four data representations including log-scaled PSD, FFT, and other statistical measures were used to perform seven-class experiments with SVM. An average accuracy of 91.7% was achieved with a combined feature vector; however, for gravel, poor accuracy of only 66.83% was obtained. Further to this, Weiss et al. tested the same method using a RWI ATRV-Jr robot and demonstrated that the SVM-based method outperformed other methods, including Brooks and Iagnemma's method, PNN, k NN, Naïve Bayes, and J4.8 decision tree [2]. In addition, a combination of vision-based and vibration-based methods was reported to significantly improve the classification rates when compared to single sensor-based prediction performance [16]. Finally, Brooks and Iagnemma introduced a self-supervised method suitable for environments with unexpected appearance [5].

Although comparatively little research has been reported on acoustic-based methods, it is a growing area of interest. Ojeda et al. used a microphone to classify robot-terrain interactions; however, the study reported that sound performed poorly as the sole modality for terrain classification, except grass [17]. Libby and Stentz employed an acoustic-based method to classify interactions with three terrain classes and three hazardous objects using a combined feature vector formed by several traditional features including spectral coefficients, moments, and temporal and spectral characteristics [11]. By applying a smoothing technique, an accuracy of 92% was obtained. More recently, Valada et al. performed a novel deep learning-based terrain classification method [18]. Using a convolutional neural network to learn deep features from spectrograms of the acoustic signal, the method was reported to significantly outperform methods using traditional audio features. Christie and Kottege presented a real-time acoustics-based terrain classification system for legged robot [8]. Extracting a 32-dimensional feature vector by combining spectral and temporal features, accuracy of 92.9% was obtained, and using a noise subtraction technique to remove servo noise, it was further increased to 95.1%. Note that the acoustic-based method is susceptible to background noise; however, due to the complexity and great variation in unpredictable background noise, apart from the servo noise subtraction method, few methods for environmental noise have been reported. A general method to evaluate the effect of noise on classification rate is to test the robustness of the proposed method during exposure to different background noises at differing levels of SNR.

In this paper, a new terrain classification framework is presented to improve classification performance. There are two main contributions in our study. First, instead of extracting features from a handcrafted domain, a two-stage feature

selection method combining ReliefF and mRMR algorithms was developed to select optimal yet compact feature subsets, which takes both attributes weights and redundancy reduction into account. Moreover, the combination method is more computationally efficient than mRMR working alone. Second, by fusing the predictions from five data sources, a multiclassifier combination method was developed. The predicted class is determined by integrating prior knowledge with the current classification results. The proposed framework has demonstrated promising performance.

2. Overview of the Terrain Classification Framework

The proposed framework involves the following steps:

- (1) Data collection from tracked robot-terrain interaction
- (2) Assigning labels to the prepared terrains
- (3) Splitting the collected signal into short time windows
- (4) Extracting features from each window
- (5) Selecting optimal feature subsets using the two-stage feature selection method
- (6) Training a classifier using the optimal feature subsets
- (7) Predicting the class labels of these short windows
- (8) Determining the terrain class by fusing the predictions from each classifier based on the prior knowledge of each data source

A schematic overview of the framework is shown in Figure 1.

2.1. Data Collection and Hand Labeling. In each experiment, the tracked robot was driven over six different types of terrain: brick (S_B), asphalt (S_A), low grass (S_{LG}), firm soil (S_{FS}), gravel (S_G), and soft soil (S_{SS}), as illustrated in Figure 2. For each type of terrain, one location was considered in the experiment. A data acquisition instrument was used to collect the signal. Data was recorded and transmitted to a desktop computer via a router in a real-time manner, and further processing was performed offline. The data was converted into MAT-files and processed within the MATLAB environment. Each sequence was split into short windows of one-second duration and all short windows were assigned a specific class label.

2.2. Feature Extraction. Features were extracted from each window. FFT is the most commonly used feature extraction method for acceleration and is often used to perform transformation from time domain to frequency domain. Moreover, it is the foundation of many other features such as the MFCC and frequency characteristics. Thus, FFT was chosen as a basic feature candidate for both the acoustic and acceleration data. As the SNR is higher in regions where there is more power, only the lower part of the spectrum was used in this study. The truncation point was set to 200 Hz.

MFCCs are perceptually based spectral features that have been successfully used in speech recognition, which basically

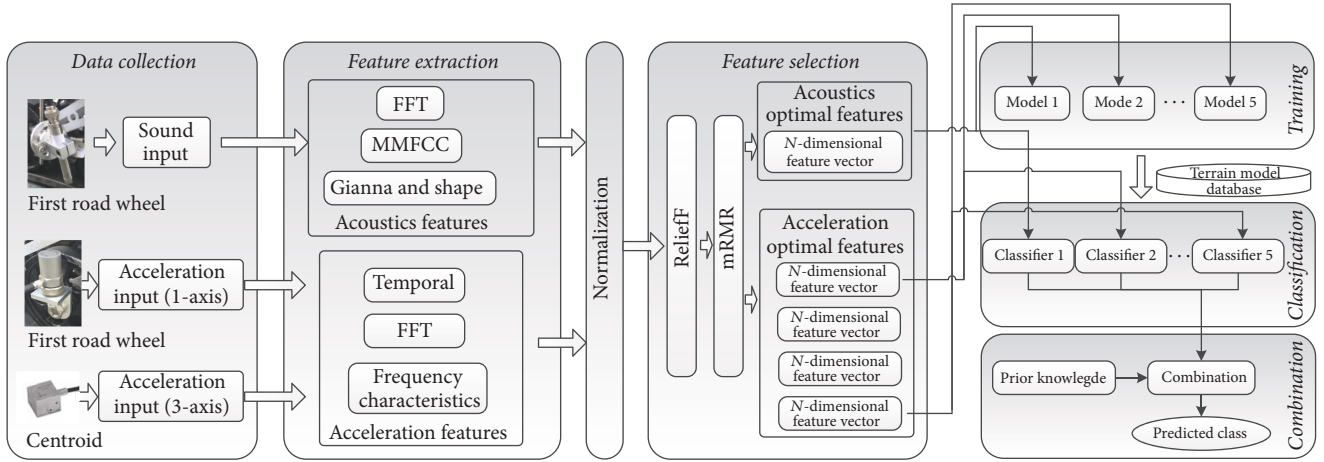


FIGURE 1: A flowchart of the proposed terrain classification framework.



FIGURE 2: Images of the different terrains used in the experiment.

maps the linear frequency scale to a scale that resembles the frequency resolution of the human ear [19]. Whereas the majority of spectral power of the acoustic signal resulting from robot-terrain interaction ranges between 0 and 200 Hz, the speech signal has more distinctive information in the mid-to-high frequency ranges; therefore, a modified form with greater emphasis on lower frequencies should be developed for this application. The traditional relationship between Mel-frequency and normal frequency can be written as

$$F_{\text{Mel}} = 2595 \lg \left(1 + \frac{f}{700} \right) \quad f = 0, \dots, 4000 \text{ Hz}, \quad (1)$$

where F_{Mel} represents the transformed frequency with the unit Mel, f represents the linear frequency, and 4000 Hz is the Nyquist frequency.

Transforming (1) to a more general form, it becomes

$$F_{\text{Mel}} = B * \lg \left(1 + \frac{f}{g} \right), \quad (2)$$

where $B = 2595$ and $g = 700$. Substituting $F_{\text{Mel}} = 2146$ Mel and $f = 4000$ Hz into (2) yields

$$2146 = B * \lg \left(1 + \frac{4000}{g} \right). \quad (3)$$

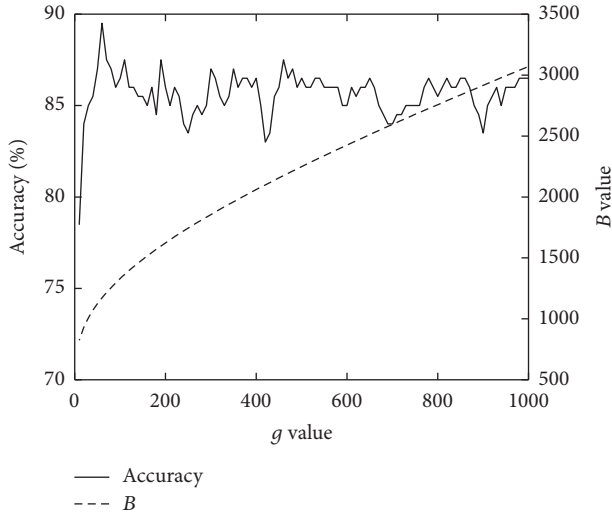


FIGURE 3: The classification accuracy as a function of g and B . The SVM and 18-dimensional MMFCCs were used to perform the classification.

The best Mel mapping function curve can be obtained by tuning B and g . According to (3), we changed the parameter g from 10 to 1000 with an interval of 10 and then 100 parameter sets (g, B) were obtained. A tuned SVM classifier was used to perform the classification with the MFCC feature vector. Finally, the parameter set that delivered the best classification performance was identified and used as the MFCC. The classification accuracy as a function of g and B is shown in Figure 3. It can be observed that best performance is achieved when $B = 1172$ and $g = 60$. In this paper, 18-dimensional modified MFCCs were used and denoted as 18-dimensional MMFCCs.

The gianna and shape feature vector [11] is a combination of the gianna feature [20] and shape feature [21]. The gianna feature includes three features extracted from the time domain (ZCR, STE, and entropy) and three features from the frequency domain (spectral centroid, spectral roll-off, and spectral flux). The ZCR indicates the number of times the signal crosses the zero axis per second. The STE is the sum of squares of the amplitudes. Abrupt energy changes in the signal are characterized by entropy. The entropy is calculated by $U = -\sum_{i=0}^{n-1} \sigma^2 \cdot \log_2(\sigma^2)$, where n denotes the number of subframes that a signal frame is divided into and σ denotes the normalized energy of a subframe. In this work, a tuned SVM classifier was used to perform the classification with the entropy feature extracted using $n = 1, 2, \dots, 20$. Finally, the value of $n = 10$ was found to give the best performance and subsequently was used as the entropy feature. Spectral roll-off is a frequency value and a certain percentage of energy lies under it. The percentage was empirically set to be 80%. Spectral flux is used to measure local spectral changes between successive frames and can be calculated by summing and squaring the differences between successive frames. The shape feature consists of four scalar features characterizing the shape of the distribution: spectral centroid, spectral standard deviation, spectral skewness, and kurtosis.

Compared to an acoustic signal, there are fewer feature vectors for acceleration. As mentioned previously, FFT was adopted as the basic feature candidate, and the truncation point was set to 200 Hz. By doing so, the lower part of the spectrum is used, resulting in a higher SNR. Based on FFT, five derived frequency domain indicators were calculated: MSE, RMSE, FC, RVE, and VF [22]. Eleven temporal indicators were also extracted, including the average value, maximum value, and RMS [23].

To summarize, each short window was taken as a dataset and processed to extract different feature vectors. The feature vector candidates for the acoustic set are listed as follows:

- (i) 18-Dimensional MMFCCs
- (ii) 200-Dimensional FFT
- (iii) 9-Dimensional gianna and shape feature
- (iv) 227-Dimensional feature vector formed by the above three feature vectors

The feature vector candidates for acceleration set are listed as follows:

- (i) 11-Dimensional temporal indicators
- (ii) 200-Dimensional FFT
- (iii) 5-Dimensional frequency characteristics
- (iv) 216-Dimensional feature vector formed by the above three feature vectors

All features were normalized to range within $[-1, 1]$ before feature selection and classification. Note that all the codes were implemented in MATLAB environment.

2.3. Feature Selection. More features can increase computation load and are also more likely to cause overfitting of the corresponding classification model. In classification problems, there are hundreds of potential features that can be used to characterize a target object; however, noisy irrelevant features provide little information and should be found and removed. In this paper, a two-stage feature selection method is developed combining ReliefF [24] and mRMR [25]. The ReliefF algorithm determines the quality attributes when there is dependency between attributes, while the mRMR algorithm finds attributes that hold the highest relevance with the target class and show maximal differentiation from each other. Since the ReliefF algorithm is less computationally expensive than the mRMR, it is performed in the first stage to select feature candidates as well as reduce computation load in the next stage. An importance value for each attribute is calculated to rank them, and the top 100 important attributes are selected to form a feature vector candidate. In the second stage, the mRMR algorithm is performed to reduce the redundancy of the feature candidates, in which the 100 selected attributes are ranked once again and can be selected based on the new ranking. Consequently, optimal yet compact feature subsets can be obtained to feed into the classification model.

The ReliefF algorithm works as follows: first, an instance R_i is randomly selected from the training sets; then search l

```

Input: feature vectors and class labels
Output: the vector  $W$  of estimations of the importance values of attributes
set  $W[A] := 0$ ;
For  $i := 1$  to  $m$  do
  randomly select a set  $R_i$ ;
  search  $l$  nearest hits  $H_j$ ;
  for each class  $C \neq \text{class}(R_i)$  do
    find  $l$  nearest misses  $M_j(C)$  from class  $C$ ;
  end
  for  $A := 1$  to  $a$  do

$$W[A] := W[A] - \frac{\sum_{j=1}^l \text{diff}(A, R_i, H_j)}{(m \cdot l)} + \frac{\sum_{C \neq \text{class}(R_i)} [(P(C)/(1 - P(\text{class}(R_i)))) \sum_{j=1}^l \text{diff}(A, R_i, M_j(C))]}{(m \cdot l)}$$

  end
end

```

ALGORITHM 1: ReliefF algorithm.

nearest neighbors from the same class and each of the different classes, which can be called nearest hits H_j and nearest misses $M_j(C)$. Based on R_i , H_j , and $M_j(C)$, the importance estimation $W[A]$ can be updated, where A denotes the set of all attributes. The pseudocode is given in Algorithm 1.

Referring to the pseudocode, $P(C)$ is the prior probability of class C , estimated from the training sets. Function $\text{diff}(A, I_1, I_2)$ calculates the difference between the values of the attribute A for two instances, I_1 and I_2 . The variable a is the number of attributes in the feature vector. In this paper, $l = 10$ and $m = 20$; both parameters are set by the users.

The mRMR algorithm employs Max-Relevance and Min-Redundancy criteria. Max-Relevance aims to find features satisfying

$$\begin{aligned} \max \quad & D(S, C), \\ D = \frac{1}{|S|} \sum_{x_i \in S} & I(x_i; C), \end{aligned} \quad (4)$$

where $I(x_i; C)$ is the mutual information value, x_i is an individual feature, and C is the class label. D is the mean value of all mutual information values.

Features selected using Max-Relevance criteria could have rich redundancy; therefore, Min-Redundancy is performed in the next step. The Min-Redundancy condition can be expressed by

$$\begin{aligned} \min \quad & R(S), \\ R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} & I(x_i; x_j). \end{aligned} \quad (5)$$

Finally, the two criteria can be combined by the operator $\Phi(D, R)$.

$$\begin{aligned} \max \quad & \Phi(D, R) \\ \Phi = & D - R. \end{aligned} \quad (6)$$

2.4. Classification. For comparison and experimental validation, four conceptually different classifiers were used to perform the classification. The principles of each classifier are briefly described as follows.

(1) *k-Nearest Neighbors.* The k NN algorithm is perhaps the simplest and most intuitive classifier [26]. It maintains good performance when there are extremely irregular decision boundaries. It is a nonparametric identification method that does not need prior probability and conditional probability density functions; however, it is labor-intensive for problems with large training sets. When classifying an unknown instance, the distance of feature vector \mathbf{X} to all training vectors is calculated and k training vectors closest to \mathbf{X} are selected. The predicted class is then defined as the most frequently occurring class among the k training vectors. Euclidean distance was utilized in this work and for feature vectors, $\mathbf{X}_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $\mathbf{X}_2 = (x_{21}, x_{22}, \dots, x_{2n})$, the Euclidean distance takes the form

$$\text{dist}(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}. \quad (7)$$

(2) *Naïve Bayes Classifier.* Due to fast and easy implementation, the Naïve Bayes classifier is widely used in pattern classification problems [27]. The basic idea is to classify an instance \mathbf{X} to a class C_i that has the highest probability considering its corresponding attributes A_i . It can be expressed as

$$C(\mathbf{X}) = \arg \max C_i. \quad (8)$$

(3) *Support Vector Machine.* The basic idea of SVM [28] is to construct a hyperplane to maximize the margin between the closest points of different classes. The method is less prone to overfitting than other methods and therefore performs particularly well for classification problems with

small datasets. Two parameters need to be tuned for the SVM model, G and γ . The parameter G is used to control the effects of errors. Higher values of G lead to a more severe effect. The data is mapped into higher-dimensional space using kernel functions, and, consequently, a nonlinear hyperplane is constructed to separate the data linearly. In this paper, the radial basis function $K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2/2\gamma^2)$ was used. The parameters G and γ were tuned using the grid search and cross validation methods. The SVM algorithm was implemented in the LIBSVM software package [29].

(4) *Random Forests*. Random forest can tackle problems with high-dimensional feature vectors [30]. It is a combination of tree classifiers. Each tree classifier is built by using a random vector sampled independently from the input vector. Finally, a unit vote is obtained from each tree classifier and the most frequently occurring class label is determined to classify the input vector. A random forest consists of N tree predictors, where N can be set to any value; N was empirically set to 500.

2.5. *Multiclassifier Combination*. As different sensors can provide complementary information, a multiclassifier combination method can be developed to improve classification performance. Voting principles are perhaps the most general and useful multiclassifier combination methods, aiming to make a consensus by fusing opinions from individual classifiers [31]. The voting principle can be expressed as follows.

Assume that there is a pattern space P that consists of M mutually exclusive sets: $P = C_1 \cup C_2 \cup \dots \cup C_M$, for $C_{i2} \forall i \in \Lambda = \{1, 2, \dots, M\}$ representing classes. In this case, six different terrains are presented; thus, $M = 6$. Assume that x is a sample from P ; a classifier (denoted by e) is used to assign a label j to x , $j \in \Lambda$, which means that x belongs to class C_j . The above event can be denoted by $e(x) = j$. Assuming V individual classifiers e_v , $v = 1, \dots, V$, an input sample x is assigned a label j_v by each classifier; then an event $e_v(x) = j_v$ is produced. Events are used to establish an integrated classifier E , and, finally, a definitive label j is given to the input sample x ; namely, $E(x) = j$, $j \in \Lambda$.

The event $e_v(x) = i$ can be expressed as

$$T_v(x \in C_i) = \begin{cases} 1, & \text{when } e_v(x) = i, i \in \Lambda \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

A commonly used voting rule based on majority is given by

$$E(x) = j, \quad (10)$$

$$\text{if } T_E(x \in C_j) = \max_{i \in \Lambda} T_E(x \in C_i) \geq \alpha * V,$$

where $0 < \alpha \leq 1$.

In this study, the confusion matrix [32] was utilized to describe the errors for each classifier e_v as follows:

$$PT_v = \begin{pmatrix} n_{11}^{(v)} & n_{12}^{(v)} & \dots & n_{1M}^{(v)} \\ n_{21}^{(v)} & n_{22}^{(v)} & & \\ \vdots & \vdots & \ddots & \\ n_{M1}^{(v)} & n_{M2}^{(v)} & \dots & n_{MM}^{(v)} \end{pmatrix}, \quad (11)$$

where $v = 1, 2, \dots, V$. $n_{ij}^{(v)}$ denotes that $n_{ij}^{(v)}$ samples belonging to class C_i are classified into class j by classifier e_v , where i denotes class C_i and j denotes the event $e_v(x) = j$.

The number of test samples is given by

$$N^{(v)} = \sum_{i=1}^M \sum_{j=1}^M n_{ij}^{(v)}. \quad (12)$$

The number of samples belonging to each class C_i is

$$n_i^{(v)} = \sum_{j=1}^M n_{ij}^{(v)} \quad i = 1, \dots, M. \quad (13)$$

The number of samples that are classified into class j is given by

$$n_{,j}^{(v)} = \sum_{i=1}^M n_{ij}^{(v)} \quad j = 1, \dots, M. \quad (14)$$

Under the occurrence $e_v(x) = j$, the probability that a sample comes from class C_i can be given by

$$P(x \in C_i | e_v(x) = j) = \frac{n_{ij}^{(v)}}{n_{,j}^{(v)}} = \frac{n_{ij}^{(v)}}{\sum_{i=1}^M n_{ij}^{(v)}} \quad (15)$$

$$j = 1, 2, \dots, M.$$

The confusion matrix is believed to be able to reflect the performance of e_v . A sample confusion matrix is shown in Figure 4. In this paper, the confusion matrix was employed as the prior knowledge from each data source, and each dataset corresponds to an individual classifier; thus, the total number of votes for class C_i is given by

$$P_E(x \in C_i) = \sum_{v=1}^V P(x \in C_i | e_v(x) = j_v) \quad (16)$$

$$i = 1, 2, \dots, M.$$

Finally, the voting principle based on prior knowledge takes the form

$$E(x) = j \quad (17)$$

$$\text{when } \exists j \in \Lambda P_E(x \in C_j) = \max_{i \in \Lambda} P_E(x \in C_i).$$

| | | | | | | |
|----------|-------|-------|----------|----------|-------|----------|
| S_B | 0.90 | 0.00 | 0.00 | 0.07 | 0.03 | 0.00 |
| S_A | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| S_{LG} | 0.00 | 0.00 | 0.85 | 0.00 | 0.07 | 0.07 |
| S_{FS} | 0.10 | 0.00 | 0.00 | 0.88 | 0.03 | 0.00 |
| S_G | 0.00 | 0.00 | 0.17 | 0.00 | 0.78 | 0.05 |
| S_{SS} | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.97 |
| | S_B | S_A | S_{LG} | S_{FS} | S_G | S_{SS} |

FIGURE 4: Confusion matrix for six-class terrain classification using SVM classifier and 18-dimensional MMFCC at speed of 2 m/s. The terrain representations: S_B : brick, S_A : asphalt, S_{LG} : low grass, S_{FS} : firm soil, S_G : gravel, and S_{SS} : soft soil.

3. Experimental System

A data acquisition system was proposed to record signals, which consisted of a data acquisition instrument (24-bit), router, computer, and three different sensors, as shown in Figure 5.

The acoustic signal was measured using an acoustic pressure sensor placed close to the first road wheel. To reduce the influence of background noise, it was pointed downward perpendicular to ground surface. For shock absorption, a bracket and damping foam block were used. A single-axis accelerometer was mounted on the axis of the first road wheel to collect acceleration data along vertical direction. Vibrations induced in the centroid position of the robot were measured by a triaxial accelerometer along three perpendicular directions, as depicted by the coordinate system shown in Figure 6. It was mounted on the bottom of robot. As shown in Figure 6, the instrumented robot has dimension of $1.3 \text{ m} \times 0.75 \text{ m} \times 0.38 \text{ m}$ and weight of approximately 75 kg. Similar studies have demonstrated that driving speeds can significantly affect the generation of sound and vibration. Here, a wireless joystick was used to roughly maintain control of the speed and 3 different speeds were considered: 0.4 m/s, 0.8 m/s, and 2 m/s. In the experiments, the sampling rates for the acoustic and acceleration signals were set to 8 kHz and 1 kHz, respectively.

To preserve symmetry, 80 samples were chosen for each class, such that a total of 480 samples were obtained. Prepared samples were separated into two sets of equal numbers, namely, the training sets and testing sets. Representative examples of the signal token from five data sources at 2 m/s are presented in Figure 7. The horizontal axis on top represents the six different types of terrains. Each terrain type was sampled for a duration of 2 seconds and is shown separated by a dotted line.

4. Results and Discussion

The results of each trial generated a confusion matrix as depicted in Figure 4. For compactness, each confusion matrix

is condensed into a single number, called the accuracy [33]. The accuracy can be calculated by averaging the true positive rates across six classes and, in this paper, it was employed as the measure of performance. Results obtained using the individual data source and handcrafted feature vector at 2 m/s are listed in Table 1.

For the acoustic-based method, accuracies between 53.8% and 89.6% were achieved. The best result in terms of accuracy was obtained using MMFCC and SVM, whereas the worst result was observed using gianna and shape and kNN. Interestingly, the results obtained for gianna and shape in this study are contrary to the conclusions reported in [11], in which a wheeled robot was used. In this study, the structural noise resulting from the track system vibration is believed to be the cause of these differences because structural noise can behave differently to the robot-terrain interaction sound. On the other hand, a wheeled robot would not produce as much structural noise as a tracked robot. Therefore, we believe it is reasonable that the gianna and shape features did not perform as well in the previous paper. Nonetheless, structural noise can also be considered as part of the robot-terrain interaction sound because the terrain being traversed causes track system vibrations. Considering all the classifiers, MMFCC outperforms the other two feature vectors. Using the acceleration-based method, FFT was better than the other two feature vectors in terms of performance. The best result was given by CX data source, FFT, and SVM. Due to poor performance, temporal and frequency characteristics features were removed from further experiments.

Referring to Table 1, for each classifier, the confusion matrices corresponding to the highest accuracy of each data source are employed as prior knowledge, as shown in Figures 4 and 8. It can be observed that different data sources have different distributions for the six terrains. Consequently, it is reasonable to believe that different data sources can complement each other and improve classification performance.

Figure 9 shows classification accuracies obtained using the proposed framework as a function of the number of optimal features. Based on the results, better performance is more likely to be achieved at higher speeds. When travelling at higher speeds, the robot can generate stronger interactions with the terrain, which results in acoustic and acceleration signals of higher magnitudes. The signals in turn lead to clearer terrain signatures. In addition, since the window size is constant for different speeds, a higher speed leads to a longer travel distance such that a larger amount of information is captured. Representative signal tokens at different speeds are shown in Figure 10.

From the classifier standpoint, SVM and RF clearly outperformed the other two simple classifiers; however, it should be pointed out that they are relatively computationally expensive. The highest accuracy of 99.6% was achieved using RF with 90 and 100 optimal features at 2 m/s. Additionally, accuracies given by RF at 0.8 m/s were above 95%, except for the trial with 10 optimal features. However, when traversing at 0.4 m/s, all the accuracies given by SVM were beyond 85% except for the trial with 10 optimal features, while most of the accuracies given by RF are lower than 85%. The worst performance was given by kNN with no accuracy

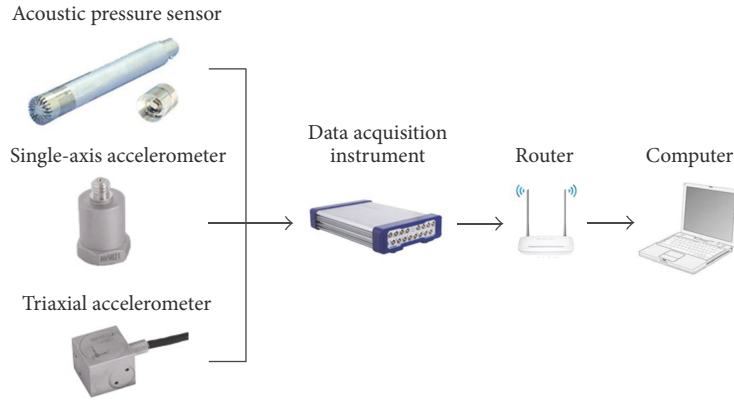


FIGURE 5: Frame diagram of the proposed data acquisition system.

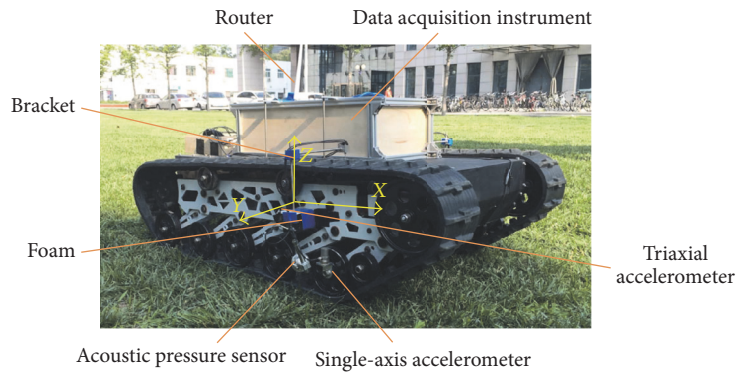


FIGURE 6: The instrumented tracked robot.

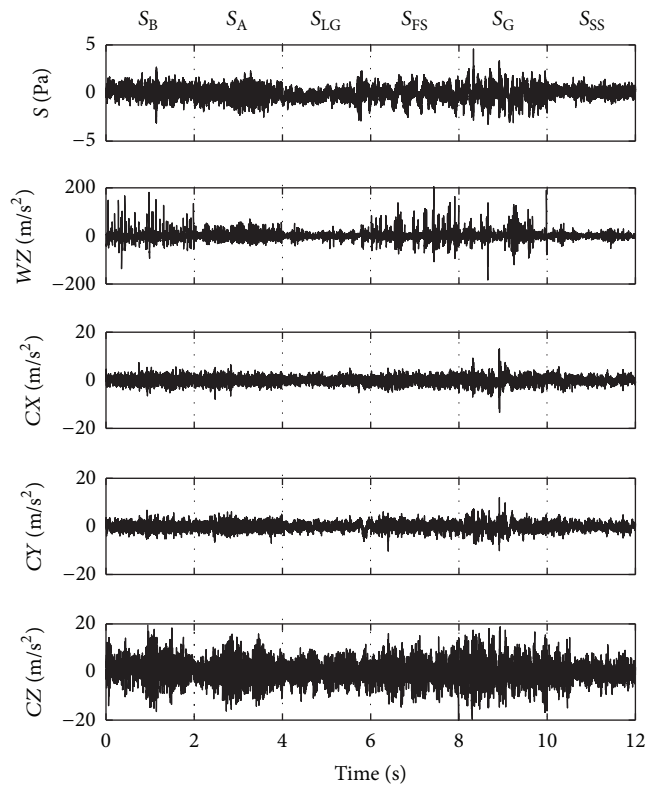


FIGURE 7: Signal token examples of five data sources at 2 m/s. S represents acoustic data source, WZ represents acceleration data source in first road wheel, CX represents X-axes (front-back) acceleration data source, CY represents Y-axes (left-right) acceleration data source, and CZ represents Z-axes (up-down) acceleration data source.

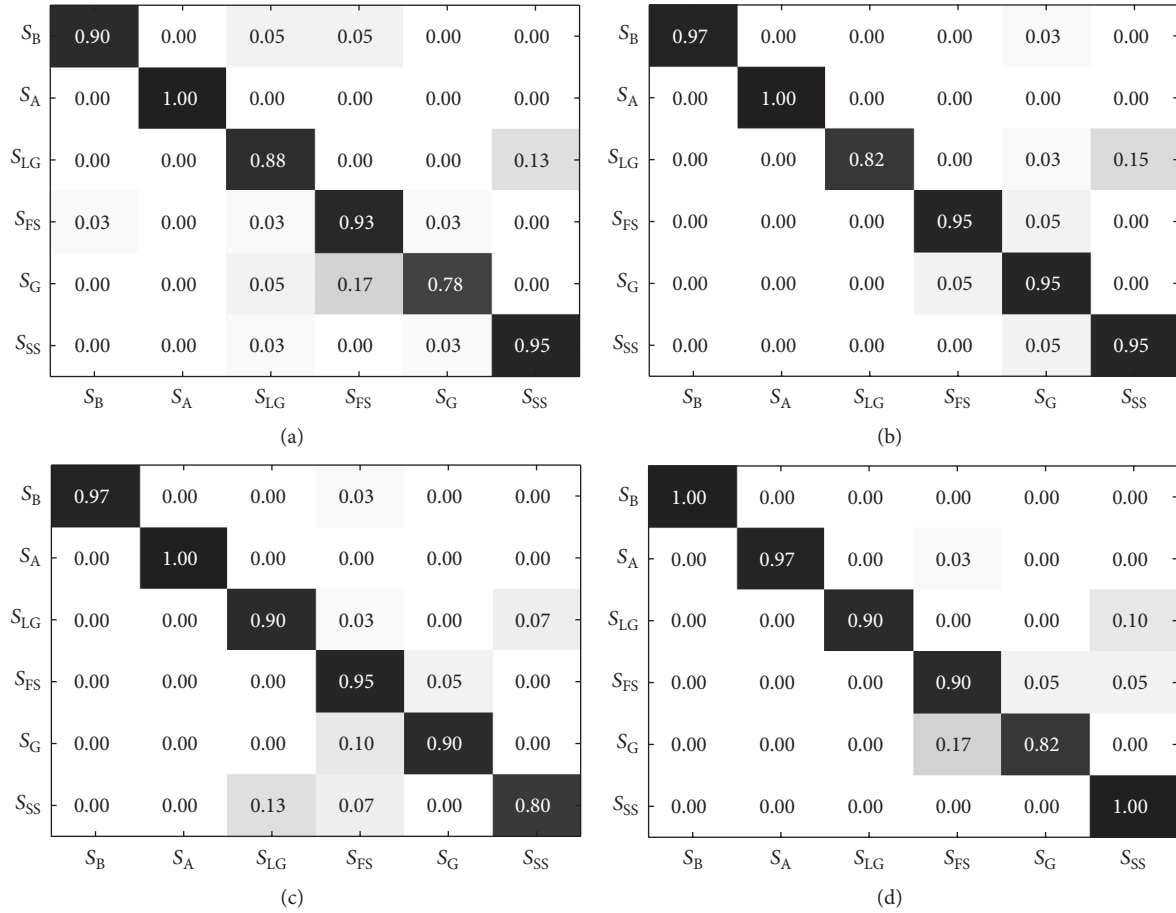


FIGURE 8: The confusion matrices obtained with SVM at 2 m/s: (a) confusion matrix of WZ with FFT and SVM; (b) confusion matrix of CX with FFT and SVM; (c) confusion matrix of CY with FFT and SVM; and (d) confusion matrix of CZ with FFT and SVM. They are used as prior knowledge for each data source when the speed is 2 m/s and SVM is used.

beyond 95%. Moreover, most of the accuracies at 0.4 m/s drop below 80%. The results given by NB presented the least fluctuation in response to the number of optimal features. The ranges of accuracies at the three different speeds were 95.4%~97.5%, 87.1%~91.3%, and 77.5%~82.5%. The margins between the accuracies at different speeds are large and very clear. In conclusion, more optimal features do not always give better accuracy, which indicates that the most useful features should be determined in order to improve classification performance.

Tables 2–4 show the degrees of improvements made using the proposed framework. Based on the excellent performance values given by MMFCC and FFT, in the future they will be adopted as the benchmark feature vectors for acoustic and vibration-based methods, respectively. In this study, the classification performance of all classifiers was improved. Moreover, greater improvements were achieved for the robot traversing at lower speeds. As explained previously, signals collected at lower speeds have lower magnitudes, and, consequently, the corresponding terrain signatures are more likely to be affected by noise. Nonetheless, in comparison with the traditional methods, our proposed terrain classification

framework succeeded in digging out discriminative information hidden in weak signals.

The purpose of terrain classification is to improve robot control and thus, in addition to classification accuracy, the classification time is another important factor used to guarantee real-time implementation. Generally speaking, algorithms resulting in faster classification times are believed to be better for running online. Figure 11 compares the classification times of the four different classifiers on a single sample, measured on a computer with an Intel Core i5-6300HQ CPU and 3.89 GB RAM. The classification times are the average values of 10 runs for each classifier. It can be observed that classification times increase approximately linearly with the number of optimal features. The classification times obtained with NB increase more rapidly with optimal features number than for other classifiers, followed by k NN, SVM, and RF. From the standpoint of classification time, this study suggests that SVM is the best approach. However, the training times required by SVM can be up to several minutes, because the grid search method used to tune the parameters is quite time-consuming. Although the training can be done offline and is less important than classification

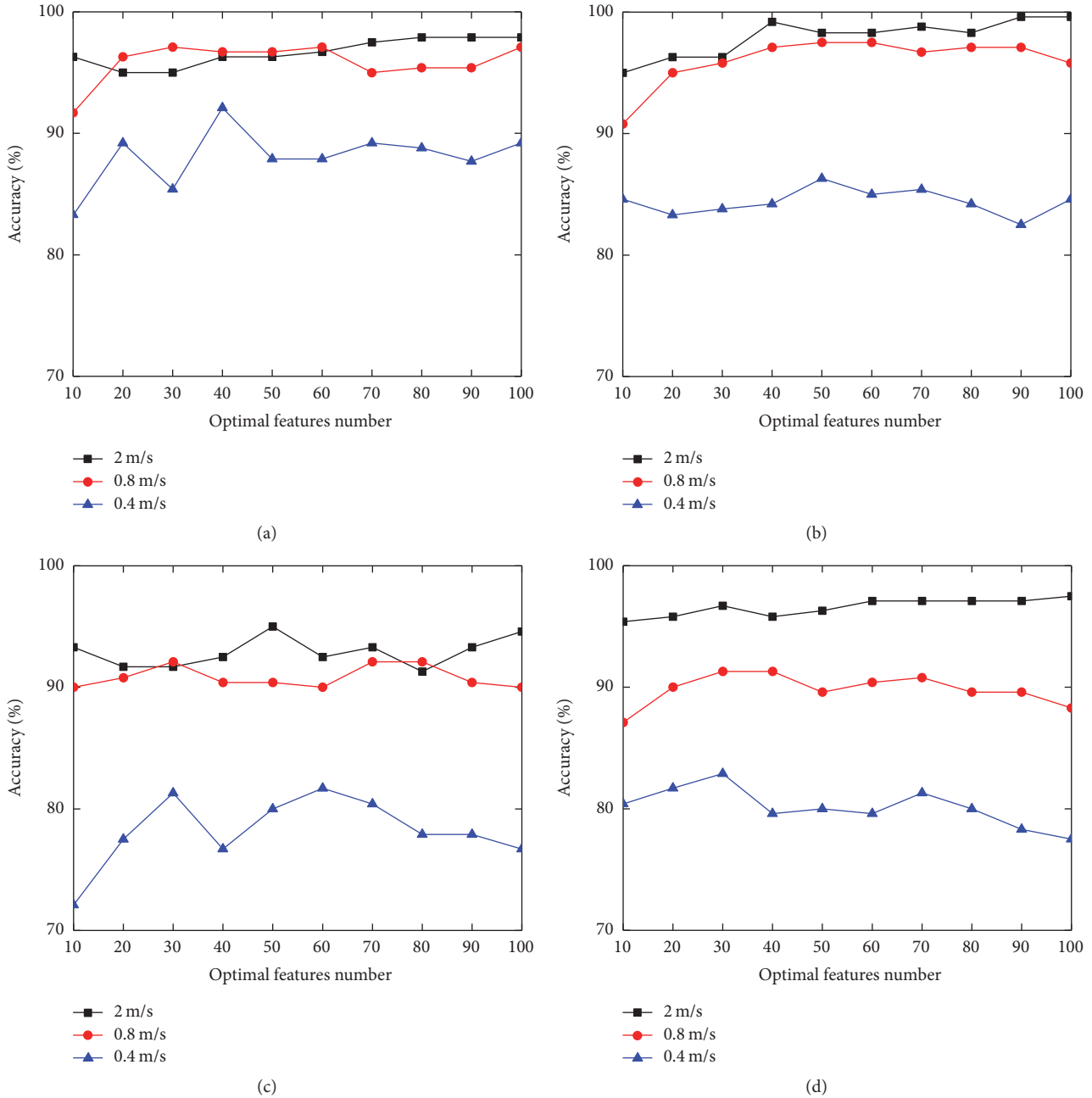


FIGURE 9: Accuracies obtained using the proposed terrain classification framework as a function of number of optimal features: (a) Accuracies of SVM; (b) accuracies of RF; (c) accuracies of k NN; and (d) accuracies of NB. The horizontal axis indicates the number of selected optimal features (dimension of optimal feature vector). The accuracies were measured for three different driving speeds.

time, it should also be considered. For larger training sets, the hours or even days of training time required by SVM would be unacceptable. While RF also suffers from long training times, it is hundredfold faster than SVM during the training phase but hundredfold slower than NB. Differing from other classifiers, k NN must use each training sample for online classifications and thus its classification time largely depends on the size of training set. For large training sets, k NN can become too slow to implement online. When there are less than 30 optimal features, NB performs better than

k NN and RF in terms of classification time. In general, it is hard to determine which algorithm is the best in terms of computation time, because it is affected by several factors such as the size of the training set, optimal features number, and offline training time.

5. Conclusions

In this paper, a new terrain classification framework was presented. The experiments were carried out with a tracked robot

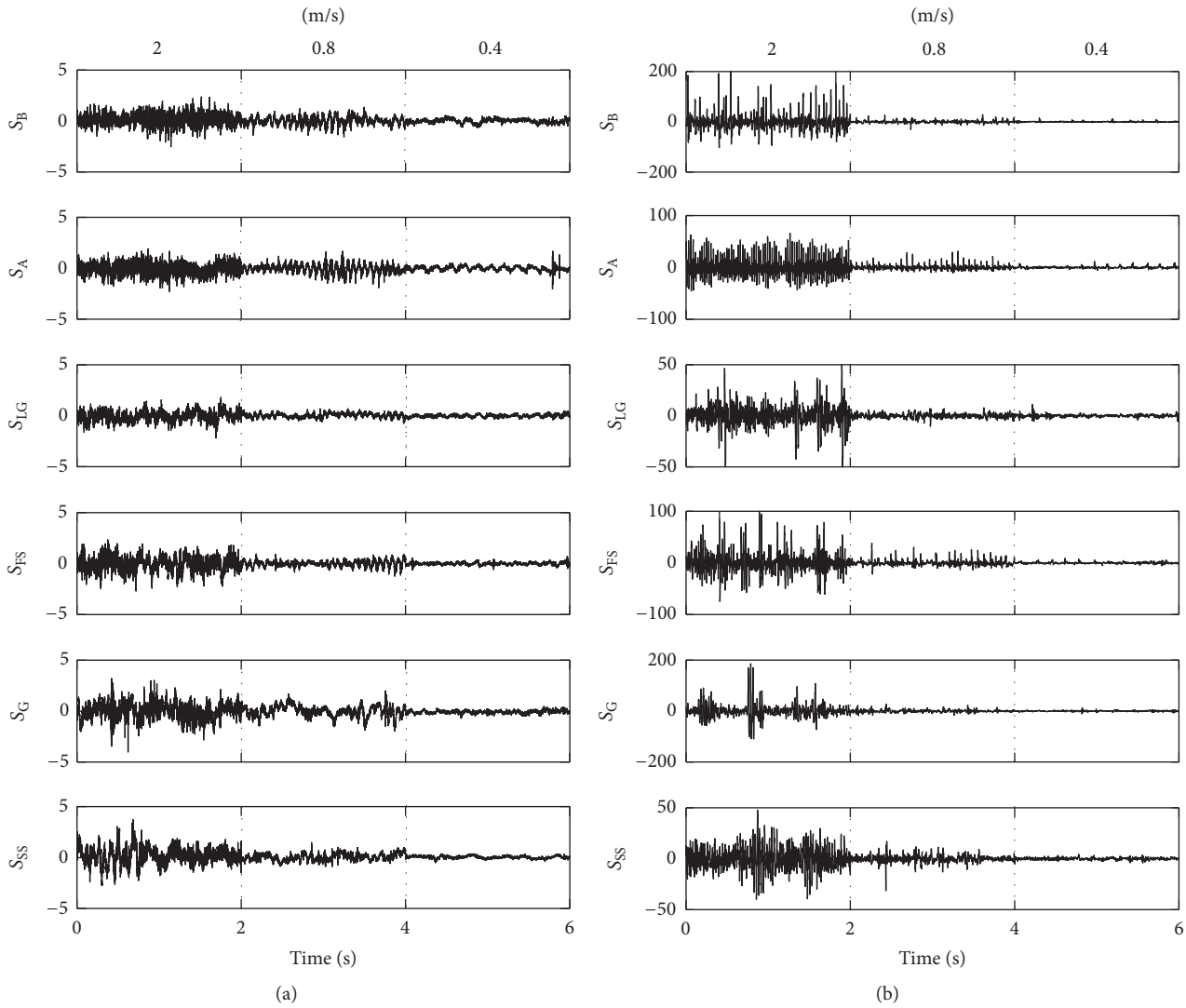


FIGURE 10: Representative signal tokens at three different speeds from (a) data source of S and (b) data source of WZ.

on six different terrains. Multiple sensors were employed to collect signals, and in total five data sources were used. A two-stage feature selection method was proposed to obtain optimal feature subsets, and a multiclassifier combination method considering prior knowledge was developed. Finally, four conceptually different classifiers were employed to perform the classification.

The results showed that the new framework successfully improves classification performance with optimal feature subsets when different classifiers are used. Only a small number of features effectively contribute to classification, which demonstrates the necessity of the feature selection operation. Different distributions of the confusion matrices resulting from five data sources revealed that complementary information can be obtained from the classifier combination. In addition, greater improvements are achieved for signals collected at lower speeds, which means that our approach can successfully dig out discriminative information hidden in the weak signals. Additionally, the accuracies tend to increase

at higher speeds, as higher speeds lead to stronger signals and longer travel distances. For real-time properties, the classification times increase approximately linearly with the number of optimal features. Since the computation time is affected by several factors such as the size of the training set, optimal features number, and offline training time, it is difficult to determine the best algorithm in relation to computation time. In this study, the SVM was found to be the best approach in terms of classification time. In comparison to traditional methods, this work suggests that the new framework could handle more complex terrain and increase the probability of detecting danger in advance due to the presence of the acoustic modality. Another advantage is that the proprioceptive sensors used in this study cost much less than tactile sensors, cameras, and LADARs. In future studies, additional types of hazards such as marshland, desert, and stream should be considered. To provide further variation, different locations should be considered within each terrain type.

TABLE 1: Accuracies with individual data source and handcrafted feature vector at 2 m/s.

| Data source | Feature vector | SVM | RF | kNN | NB |
|--------------------------------|---------------------------|-------|-------|-------|-------|
| Acoustics (S) | MMFCC | 89.6% | 87.1% | 80.8% | 77.9% |
| | FFT | 82.9% | 70% | 56.7% | 55.4% |
| | Gianna and shape | 58.8% | 67.9% | 53.8% | 64.2% |
| Wheel vibration (WZ) | Temporal | 68.3% | 62.5% | 62.1% | 56.3% |
| | FFT | 90.4% | 89.2% | 75.4% | 84.6% |
| | Frequency characteristics | 59.6% | 64.2% | 62.5% | 51.7% |
| Centroid vibration X-Axes (CX) | Temporal | 41.7% | 52.9% | 51.3% | 47.5% |
| | FFT | 94.2% | 92.9% | 72.1% | 88.3% |
| | Frequency characteristics | 46.7% | 54.2% | 47.1% | 44.2% |
| Centroid vibration Y-Axes (CY) | Temporal | 42.5% | 44.2% | 50% | 46.7% |
| | FFT | 92.1% | 93% | 91.3% | 88.2% |
| | Frequency characteristics | 44.6% | 44.6% | 46.7% | 45.1% |
| Centroid vibration Z-Axes (CZ) | Temporal | 44.2% | 46.3% | 50% | 45.5% |
| | FFT | 93.3% | 92.9% | 91.3% | 90.8% |
| | Frequency characteristics | 42.5% | 48.3% | 46.7% | 42.5% |

TABLE 2: Comparison of the proposed terrain classification framework with traditional method on benchmark feature vectors. MMFCC and FFT are adopted as benchmark feature vectors for acoustics and acceleration data sources, respectively. The driving speed is 2 m/s. The number in brackets is the dimensionality of the optimal feature subset corresponding to the best result.

| Data source | Feature vector | SVM | RF | kNN | NB |
|--------------------|----------------|-------------|-------------|-----------|--------------|
| S | MMFCC (18D) | 89.6% | 87.1% | 80.8% | 77.9% |
| WZ | FFT (200D) | 90.4% | 89.17% | 75.42% | 84.6% |
| CX | FFT (200D) | 94.2% | 92.92% | 72.08% | 88.3% |
| CY | FFT (200D) | 92.1% | 92.95% | 91.3% | 88.2% |
| CZ | FFT (200D) | 93.3% | 92.92% | 91.3% | 90.8% |
| Proposed framework | | 97.9% (80D) | 99.6% (90D) | 95% (50D) | 97.5% (100D) |

TABLE 3: Comparison of the proposed terrain classification framework with traditional method on benchmark feature vectors. MMFCC and FFT are adopted as benchmark feature vectors for acoustics and acceleration data sources, respectively. The driving speed is 0.8 m/s. The number in brackets is the dimensionality of the optimal feature subset corresponding to the best result.

| Data source | Feature vector | SVM | RF | kNN | NB |
|--------------------|----------------|-------------|-------------|-------------|-------------|
| S | MMFCC (18D) | 76.7% | 75% | 67.5% | 72.1% |
| WZ | FFT (200D) | 80% | 82.5% | 76% | 75.8% |
| CX | FFT (200D) | 82.5% | 82.9% | 76.3% | 76.3% |
| CY | FFT (200D) | 87.5% | 89.2% | 79.6% | 79.6% |
| CZ | FFT (200D) | 88.3% | 85.4% | 80% | 80% |
| Proposed framework | | 97.1% (30D) | 97.5% (50D) | 92.1% (30D) | 91.3% (30D) |

TABLE 4: Comparison of the proposed terrain classification framework with traditional method on benchmark feature vectors. MMFCC and FFT are adopted as benchmark feature vectors for acoustics and acceleration data sources, respectively. The driving speed is 0.4 m/s. The number in brackets is the dimensionality of the optimal feature subset corresponding to the best result.

| Data source | Feature vector | SVM | RF | kNN | NB |
|--------------------|----------------|-------------|-------------|-------------|-------------|
| S | MMFCC (18D) | 74.2% | 67.9% | 62.5% | 68.3% |
| WZ | FFT (200D) | 80% | 63% | 76% | 53.8% |
| CX | FFT (200D) | 69.2% | 69.2% | 76.3% | 60.8% |
| CY | FFT (200D) | 82.5% | 79.6% | 79.6% | 70% |
| CZ | FFT (200D) | 83.8% | 80% | 80% | 73.3% |
| Proposed framework | | 92.1% (40D) | 86.3% (50D) | 81.7% (60D) | 82.9% (30D) |

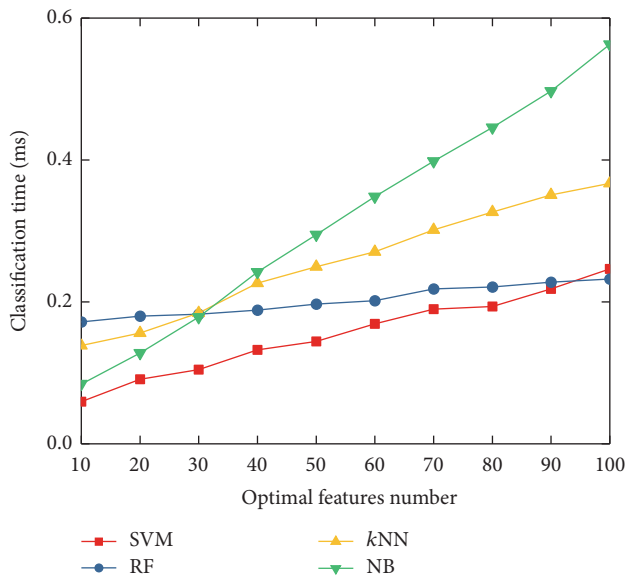


FIGURE 11: Classification times of different classifiers as a function of the number of optimal features.

Acronyms

| | |
|---------|---|
| LADAR: | LASer Detection And Ranging |
| PSD: | Power spectral density |
| PCA: | Principal component analysis |
| FFT: | Fast Fourier transform |
| SVM: | Support vector machine |
| PNN: | Probabilistic neural network |
| kNN: | k -Nearest neighbors |
| NB: | Naïve Bayes |
| SNR: | Signal-to-noise ratio |
| mRMR: | Minimal-redundancy-maximal-relevance |
| MFCC: | Mel-frequency cepstrum coefficient |
| MMFCC: | Modified Mel-frequency cepstrum coefficient |
| ZCR: | Zero crossing rate |
| STE: | Short time energy |
| MSF: | Mean square frequency |
| RMSF: | Root mean square frequency |
| FC: | Frequency center |
| RVF: | Root variance frequency |
| VF: | Variance frequency |
| RMS: | Root mean square |
| LIBSVM: | Library for support vector machines |
| RF: | Random forests. |

Conflicts of Interest

The authors declare that they have no conflicts of interest.

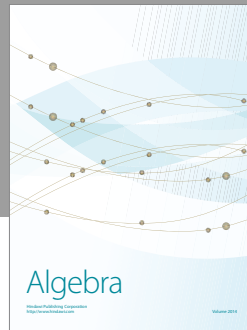
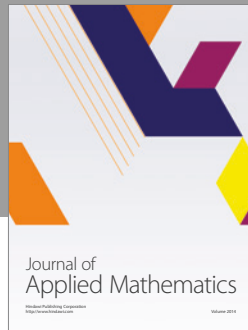
Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China (Grant no. U1564210).

References

- [1] P. Papadakis, "Terrain traversability analysis methods for unmanned ground vehicles: a survey," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1373–1385, 2013.
- [2] C. Weiss, N. Fechner, M. Stark, and A. Zell, "Comparison of different approaches to vibration-based terrain classification," in *Proceedings of the 3rd European Conference on Mobile Robots*, pp. 7–12, Freiburg, Germany, September 2007.
- [3] H. Inotsume, K. Skonieczny, and D. S. Wettergreen, "Analysis of grouser performance to develop guidelines for design for planetary rovers," in *Proceedings of the 12th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS 2014)*, pp. 1–9, Montreal, Canada, June, 2014.
- [4] M. Ono, T. J. Fuchs, A. Steffy, M. Maimone, and J. Yen, "Risk-aware planetary rover operation: Autonomous terrain classification and path planning," in *Proceedings of the 2015 IEEE Aerospace Conference, AERO 2015*, pp. 1–10, Big Sky, Mont, USA, March 2015.
- [5] C. A. Brooks and K. Iagnemma, "Self-supervised terrain classification for planetary surface exploration rovers," *Journal of Field Robotics*, vol. 29, no. 3, pp. 445–468, 2012.
- [6] H. Zhang and J. P. Ostrowski, "Visual motion planning for mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 2, pp. 199–208, 2002.
- [7] J.-F. Lalonde, N. Vandapel, D. F. Huber, and M. Hebert, "Natural terrain classification using three-dimensional lidar data for ground robot mobility," *Journal of Field Robotics*, vol. 23, no. 10, pp. 839–861, 2006.
- [8] J. Christie and N. Kottege, "Acoustics based terrain classification for legged robots," in *Proceedings of the 2016 IEEE International Conference on Robotics and Automation, ICRA 2016*, pp. 3596–3603, Stockholm, Sweden, May 2016.
- [9] Y. Qin, M. Dong, F. Zhao, R. Langari, and L. Gu, "Road profile classification for vehicle semi-active suspension system based on Adaptive Neuro-Fuzzy Inference System," in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 1533–1538, Osaka, Japan, December 2015.
- [10] Y. Qin, C. Xiang, Z. Wang, and M. Dong, "Road excitation classification for semi-active suspension system based on system response," *Journal of Vibration and Control*, 2017.
- [11] J. Libby and A. J. Stentz, "Using sound to classify vehicle-terrain interactions in outdoor environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3559–3566, St Paul, Minn, USA, 2012.
- [12] K. Walas, "Terrain classification and negotiation with a walking robot," *Journal of Intelligent & Robotic Systems*, vol. 78, no. 3-4, pp. 401–423, 2015.
- [13] K. Iagnemma, C. Brooks, and S. Dubowsky, "Visual, tactile, and vibration-based terrain analysis for planetary rovers," in *Proceedings of the 2004 IEEE Aerospace Conference*, pp. 841–848, Big Sky, Mont, USA, March 2004.
- [14] C. Brooks, K. Iagnemma, and S. Dubowsky, "Vibration-based terrain analysis for mobile robots," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 3415–3420, Barcelona, Spain, April 2005.
- [15] C. Weiss, H. Fröhlich, and A. Zell, "Vibration-based terrain classification using support vector machines," in *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006*, pp. 4429–4434, Beijing, China, October 2006.

- [16] C. Weiss, H. Tamimi, and A. Zell, "A combination of vision- and vibration-based terrain classification," in *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp. 2204–2209, Nice, France, September 2008.
- [17] L. Ojeda, J. Borenstein, G. Witus, and R. Karlsen, "Terrain characterization and classification with a mobile robot," *Journal of Field Robotics*, vol. 23, no. 2, pp. 103–122, 2006.
- [18] A. Valada, L. Spinello, and W. Burgard, "Deep feature learning for acoustics-based terrain classification," in *Proceedings of the International Symposium on Robotics Research (ISRR)*, vol. 3, pp. 1383–1399, Genova, Italy, September 2015.
- [19] S. E. Bou-Ghazale and J. H. L. Hansen, "A comparative study of traditional and newly proposed features for recognition of speech under stress," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, pp. 429–442, 2000.
- [20] T. Giannakopoulos, D. Kosmopoulos, A. Aristidou, and S. Theodoridis, "Violence content classification using audio features," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3955, pp. 502–507, 2006.
- [21] M. C. Wellman, N. Srouf, and D. B. Hillis, "Feature extraction and fusion of acoustic and seismic sensors for target identification," in *Proceedings of the Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, pp. 139–145, Orlando, Fla, USA, April 1997.
- [22] J. Tang, Z. Deng, Q. Quan, and S. Jiang, "Real-Time Drilling Strategy for Planetary Sampling: Method and Validation," *Journal of Aerospace Engineering*, vol. 29, no. 5, article 04016033, 2016.
- [23] Q. Li, K. Xue, H. Xu, W. Pan, and T. Wang, "Vibration-based terrain classification for mobile robots using support vector machine," *Robot*, vol. 34, no. 6, pp. 660–667, 2012.
- [24] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [26] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [27] J. Han and M. Kamber, *Data Mining Concept and Techniques*, Elsevier, San Francisco, Calif, USA, 2nd edition, 2006.
- [28] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [29] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [30] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [31] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.
- [32] F. Provost and R. Kohavi, "Guest editors' introduction: On applied research in machine learning," *Machine Learning*, vol. 30, no. 2-3, pp. 127–132, 1998.
- [33] A. V. Phan, M. L. Nguyen, and L. T. Bui, "Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems," *Applied Intelligence*, vol. 46, no. 2, pp. 455–469, 2017.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

