Analyzing Seismic Signal Using Support Vector Machine for Vehicle Motion Detection

Thang Duong Nhat[1(✉)] and Mai Nguyen Thi Phuong[2]

[1] Center for Training of Excellent Students,
Hanoi University of Science and Technology, Hanoi, Vietnam
thang.dn120885@sis.hust.edu.vn
[2] Department of Precision Mechanical and Optical Engineering,
Hanoi University of Science and Technology, Hanoi, Vietnam
mai.nguyenthiphuong@hust.edu.vn

Abstract. A system to process seismic signals of vehicles passing between two sensor stations had been developed and experimented. To evaluate the feasibility of the system before field test with a real vehicle and to support the classification model with artificial data later, the input seismic data were simulated from Green's method function that accounts only for Rayleigh surface wave. The system using the Machine Learning Classification method SVM to classify data collected from two stations at any time have the state of passed or not. By processing the signal, the system could detect whether the vehicle had passed the crossing line or not with the accuracy of 99.10% for *simulated* data and 94.22% for *experiment* data. The experiment and results suggested that processing seismic signals to monitor control lines is feasible.

Keywords: Machine Learning · Seismic signal · Motion detection

1 Introduction

The monitoring tasks over an area for vehicle detection have many applications, most commonly for security over an interested area. For example, a chemical weapons production facility needs to detect illegal or suspicious movements quickly, so an official can take suitable actions in response. For a wide surveillance zone, it would be costly and difficult to only use personnel to cover the area completely. For alternative technology solutions, many motion detectors are developed with unsuitable characteristic such as the Active Sensors with high energy consumption that needed power supplement regularly are impractical in many cases.

For peace-keeping and security tasks, the system is required to have a medium - large working range, can sustain for a long period of time and does not emit signal that notify any unwanted party. Thus, the author's approach turns to the seismic wave analyzing. This is a passive detection technique, so the equipment

T. Q. Duong and N.-S. Vo (Eds.): INISCOM 2018, LNICST 257, pp. 178–190, 2019.
https://doi.org/10.1007/978-3-030-05873-9_15

would not emit any signal to the environment. At the same time, it is also cheaper, smaller and has a wider range of working than some other sensors.

For the passive detection methods, a large proportion of research is based on processing acoustic signal, image and infrared signal [3,5,9]. Seismic signal, on the other hand, is being studied less because it is more complicated. It consists of different types of wave, propagates in different forms, with different speeds and directions, and are dependent on the geology of the interested environment.

Despise all the difficulty, this is still an attractive research approach for the stated problem. The reasons for that was the seismic waves are less sensitive to Doppler effects, noises introduced by the moving vehicle and atmosphere compare to sound, image, and infrared signals. Seismic wave also holds the possibility for non-line of sight detection at significant range (Table 1).

Table 1. Capabilities of the seismic detection method [10]

Target type	Detection range (m)
Vehicles-wheeled (light)	200
Vehicles-wheeled (heavy)	400
Vehicles-tracked (light)	500
Vehicles- tracked (heavy)	1000

The seismic wave is categorized into two main type: body waves (e.g. compressional (P) wave, shear (S) wave) and surface waves (e.g. Rayleigh wave, Love wave). At the measuring point, the signal collected is the sum of both body waves that propagate in three dimensions through the interior of the earth and the surface waves that propagate in two dimensions through the surface of the earth. This property tells us that the diminishing rate of a signal for surface waves is R^2, much less than R^3 of the body waves. Hence, most of the signal gather at sampling point come from surface waves.

When considering surface waves, Rayleigh wave holds the largest proportion of impact energy, 67%, while that of the shear wave and the compressional wave are 26% and 7% [8]

The seismic wave propagation to the surrounding in a spherical surface for body waves and in a circle for surface waves. Though there are multiple kinds of waves, when the faraway target are detected, only the Rayleigh wave, which accounted the most part of the energy in the seismic wave, can be relied on to analyze. Hence, in this research, the main focused were simulating and analyzing Rayleigh wave for vehicle detection.

2 Experiment

The experiment took place on the site of Hanoi University of Science and Technology, Hanoi, Vietnam in 28 Apr. 2018. The target of the experiment was to

collect seismic signal generated from the vehicle motion to test the function of the detection system.

In the experiment, the following equipments was used: two sensor stations each with a geophones LGT-20D10 and a circuit with opamp OPA2134PA for collecting data; 3 marking stations each with one motion sensor module HC-SR501 for marking the begin, the end and the threshold; two notebook computers, two Arduino Uno, battery, relay SRD-12VDC-SL-C, IC LM7805CV TO220 and connecting cables.

The experiment use the default Arduino 10-bit ADC port with internal reference voltage of 5 V. To eliminate the 50 Hz 'hum' noise from the power grid, the Arduinos are power by the Notebook's battery through the connecting cables. The circuit with opamp OPA2134PA was built to offset signal with 2.5 V base and 0 gain. The experiment setup with two Sensor stations S0, S1 and three Marking stations M0, M1, M2 is described in the figure below. When there is motion at M0 or M2, the signal collected from the geophones transmitted to the offset circuit, converted to digital and recorded in a Arduino for each geophone with Sampling frequency of 400 Hz. The M1 sensor marking the threshold would notify both Arduino when the vehicle moved past it and saved labeling data. To synchronize the starting, threshold and stopping moment between both Arduinos, each of the three Marking stations is connected to both Arduinos by connecting cables. After collected, the data were manually transfered into the notebook for later processing (Fig. 1).

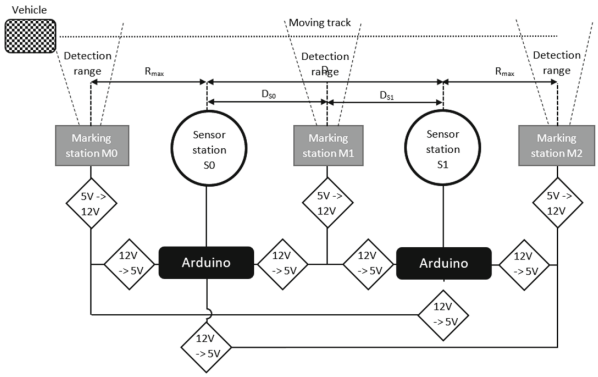

Fig. 1. Experiment setup

The geophones were placed 0.15 m underground and 0.5 m away from roadside of the testing common road made of asphalt. The Marking stations were placed at roadside with the height of 0.25 m from the road surface. The geophone testing range is R_{max} and the distance between the geophones are D. The vehicle were measured at three different speeds V, with three repetitions for each. The speeds were controlled by the driver watching the speedometer so that any acceleration

and deceleration happened outside M0-M2 range and the speed was remained constant between them. Recording was started and stopped automatically when the sensors at Marking station detected vehicle motion in the road. The vehicle used in the experiment was a KIA Morning 2015 car.

3 Method

3.1 Simulated Data Generation

Firstly, a simple quarter car model (QCM) had been used by the author to represent the forces exerted to the ground as the vehicle move over irregular surfaces. For wheeled vehicles moving over the perfectly flat ground, the irregularity forces were still presented due to the small gaps in tire treads. To justify QCM as a valid vehicle simplification, several assumptions had been made [7]:

- A point contact patch assumption is deemed sufficient as typical wavelengths of generated Rayleigh waves are greater than the characteristic dimensions of a vehicle.
- Total vehicle mass is distributed evenly to all wheel stations at all time.
- The road surface is rigid.
- Freezing the low frequency 'body bounce' vibration (around 1–2 Hz). This assumption can be made since the generated ground vibrations are usually at high frequency.

In Fig. 2, the F_t represented the force exerted by the compression of the tire spring due to the vertical displacement of the wheel z_r. Thus, the vertical displacement of the wheel could be represented by a Frequency response function (FRF) with the input $z_r(t)$.

The input $z_r(t)$ as shown in Fig. 2 is the elevation changes caused by the tire tread and the irregularity of the surface road. For simplicity, the variation in surface profile over which the wheel (modeled as a point contact) traverses could be estimated as a finite series of a half sine wave pulses.

$$\begin{cases} z_r(t) = z_{r_{max}} \sin(2\pi f_{tr} t) & for\ z_r(t) \geq 0 \\ z_r(t) = 0 & otherwise \end{cases} \tag{1}$$

The frequency of the input f_{tr} could be calculated by the corresponded moving velocity V of the vehicle over the tread pitch a. Using a simple Fourier Integration, the input in the frequency domain had the expression:

$$z_r(\omega) = \int_{-\infty}^{\infty} z_r(t)\, e^{-i\omega t} dt \tag{2}$$

According to the QCM, the displacement of the wheel from its static position are described in the equation:

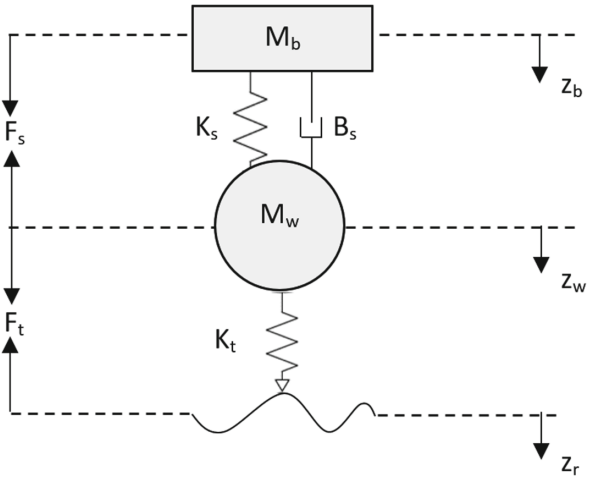

Fig. 2. A quarter car vehicle model

$$M_w \frac{\partial^2 z_w}{\partial^2 t} + B_s \frac{\partial z_w}{\partial t} + (K_t + K_s)\, z_w = K_t z_w\,(vt) \tag{3}$$

where K_t and K_s are the Tire compliance and the Suppressing spring stiffness individually. Using Fourier transform to solve Eq. (3), the elevation of the wheel in the frequency domain is:

$$z_w\,(\omega) = \frac{\omega_1^2\ z_r\,(\omega)}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + (2\omega\alpha)^2}} exp\left[-i\ tan^{-1}\left(\frac{2\omega\alpha}{\omega_0^2 - \omega^2}\right)\right] \tag{4}$$

where $\omega_0 = ((K_t + K_s)/M_w)^{1/2}$ is the Hop resonance frequency, $\omega_1 = (K_t/M_w)^{1/2}$ is the Tire "bouncing" resonance frequency and $\alpha = B_s/2M_w$ is the Normalized damping coefficient.

Having both the displacement of the wheel and the input signal in the frequency domain, the equation of the force spectrum for a single wheel axle could be established:

$$F_t\,(\omega) = K_t[(z_w\,(\omega) - z_r\,(\omega)] \tag{5}$$

The QCM described above is only valid for modeling a single axle wheel displacement. Considering the effects of multiple axles, a simple superposition of all-wheel hop displacement responses should be taken to establish the ground force spectra. The wheel hop response differently at each axle differs by a phase shift that depended on the distance from it to the front axle E_{1n} divided by the vehicle forward speed V

$$F_z^{mw}(\omega) = F_z(\omega).\left(1 + \exp\left(i\omega\frac{E_{12}}{V}\right) + \exp\left(i\omega\frac{E_{13}}{V}\right)\ldots + \exp\left(i\omega\frac{E_{1N}}{V}\right)\right) \quad (6)$$

where $F_z(\omega)$ is the force spectrum for a single wheel axle.

Another thing needed to be calculate is the Rayleigh determinant and its derivative:

$$F(k) = \left(2k^2 - k_s^2\right)^2 - 4k^2 v_l v_s \quad (7)$$

where k is the projected distance onto the $z = 0$ plane of the current wave vector, $v_l, s = (k^2 - k_l, s^2)^{1/2}$ are no specified expressions, $k_{l,s} = \omega/c_{l,s}$ are the wavenumber of bulk longitudinal and shear acoustic waves. $c_l = [(\lambda + 2\mu)/\rho]^{1/2}$ is the phase velocities of bulk longitudinal acoustic waves, $c_s = (\mu/\rho)^{1/2}$ is the phase velocities of shear acoustic waves, where $\lambda = 2\mu\sigma/(1-\sigma)$ is the Lame first parameter, σ is the Poisson's ratio, μ is the shear modulus and ρ is the soil mass density.

Note that in the consideration case, the contribution of bulk waves to the ground vibration field generated on the surface are proportional to $(k_l \cdot r)^{-2}$ and $(k_s \cdot r)^{-2}$ respectively for longitudinal and shear waves, where r is the distance from the vibration source to the observation point. For comparison, the Rayleigh waves contribution is proportional only to $(k_R\, r)^{-1/2}$. Thus, further calculations will take into account only the contribution of Rayleigh surface waves (in Eq. (7): $k = k_R$, where k_R is the wave number of a Rayleigh wave).

Solving equation $F(k) = 0$ [6], the k_R was achieved as the real root of this equation, thus also determined the velocity of Rayleigh waves $c_R = \omega/k_R$. Taking account of attenuation of generated ground vibrations in the ground result in $k_R = (\omega/c_R)(1 + i\gamma)$, where $0 < \gamma \ll 1$ is the Loss factor which describes the linear dependence of a Rayleigh wave attenuation coefficient on frequency ω.

Next, the vibration spectra generated by the vehicle-induced ground forces using Green's function method (taking into account only generated Rayleigh surface waves) were expressed:

$$v_z(\omega) = \left(\frac{2\pi}{k_R.r}\right)^{\frac{1}{2}} \frac{(-i\omega)\, k_R k_S v_s v_l}{2\pi\mu F'(k_R)} F_z^{mw}(\omega).\, e^{-k_R\gamma.r}.\, e^{ik_R r - \frac{3\pi}{4}} \quad (8)$$

where $F(k_R)$ is the derivative of $F(k)$ taken at $k = k_R$ and R is the distance from the vibration source to the observation point.

To develop a more robust system, a family of NON-LINEAR moving tracks, which still satisfy the requirement that the moving velocity is a constant, needed to be generated. Thus, the experiment setup was developed and described in Fig. 3.

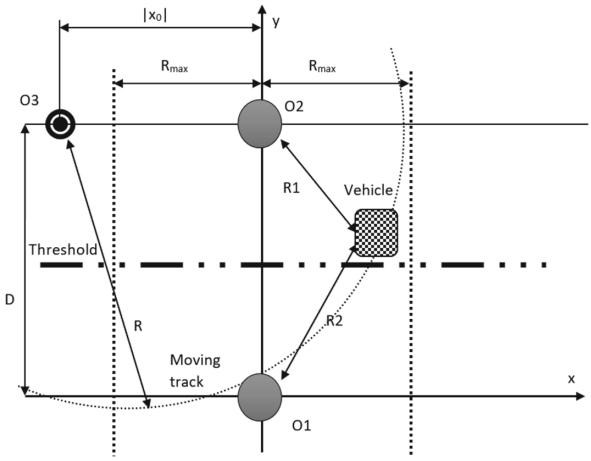

Fig. 3. Simulated experiment setup

where the two sensors were placed at O1 and O2 which have the distance between two points is D and R_{max} is the working radius of the geophone.

There were infinite number of Nonlinear moving tracks from line $y = D$ to line $y = 0$, but to satisfy the requirement of constant moving velocity, the solutions were limited than before. Thus, the moving track of vehicle was chosen as a function of a circle in Descartes coordinate system.

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \tag{9}$$

$$\begin{cases} x = x_0 + Rsin\left(\frac{V}{R}t + \varphi\right) \\ y = y_0 + Rcos\left(\frac{V}{R}t + \varphi\right) \end{cases} \tag{10}$$

To applied the simulated location r to Eq. (8), notice that the distance r from the vehicle to the observation point here is a constant. Since the vehicle was moving continuously while collecting data, the distance from the vehicle to the observation point was also varied. To use the Eq. (8), some assumption is required to simplify the problem.

The traveling distance of the vehicle were divided into n small bins. An assumption had been made that when the vehicle is inside a bin, the distance between it and the observation point remained unchanged and equaled to the distance of the center of the bin to the observation point. Thus, the Eq. (8) could be used to obtain a vertical displacement spectrum for each bin with its corresponding R_n (Fig. 4).

After acquiring a set of vertical displacement of the wheel in the frequency domain with respect to each Rn to the observation point, performed Inverse Discrete Fourier Transform (Inverse Fast Fourier Transform to be precise) and

Fig. 4. Vehicle travel distance is divided into small bins in which all data inside a bin have the same distance R_n to the observation point.

merged all data with respect to each bin defined before. The result simulated signal has the form in Fig. 5.

The parameters used in the simulation process were shown in Table 2.

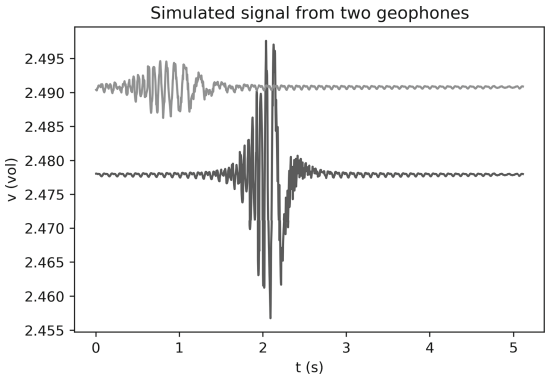

Fig. 5. Simulated signals of 2 sensors. Each peak at the time when the vehicle is closest to the sensors.

3.2 Data Classification with Support Vector Machines (SVM)

SVM is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. It was first proposed by Cortes and Vapnik [4]. In 1992, a method to create a nonlinear classifier is proposed by Boser, Guyon and Vapnik [2]. This method applied Kernel trick, originally proposed by Aizerman et al. [1], to maximum the margin hyperplane. Since there are available library that had been optimized, the Sklearn library had been chosen to implement SVM.

After acquiring the labeled data, an attempt was made to train the model directly with the Raw data by SVM algorithm. The result was very poor and highly sensitive to the skewness of the data. Looking back into Green's function method to find the vertical displacement in Eq. (8), it shown that the displacement was dependence on the distance in the Frequency Domain but not in the Time Domain. Thus, the main features to train the model was chosen to be the transformed data in the frequency domain.

Table 2. The simulated parameters

QCM parameters	Symbol/Unit	Kia Morning 2015
Total vehicle mass	M_v/kg	940
Mass of wheel	M_w/kg	∼15
Number of wheel	N_w	4
Vehicle forward velocity	V/kmh^{-1}	5, 10, 15
Tread pitch	a/m	1.409^{-3}
Magnitude of discontinuity	$z_{r_{max}}$/m	0.005
Wheelbases	$E_1 2$/m	2.385
Soil mass density	ρ/kgm^{-3}	1800
Shear modulus	μ/Nm^{-2}	$4x10^7$
Loss factor	γ	0.05
Poisson's ratio	σ	0.25
Geophone testing range	R_{max}/m	10
Distance between geophones	D/m	20
Distance between threshold to S0	D_{S0}/m	10
Distance between threshold to S1	D_{S1}/m	10

The hyperparameter *window size* was chosen to represent the number of data taken to analyzed at each individual sampling moment. As a common practice, the Discrete Fourier Transform (DFT) was not used directly to convert data from the time domain to the frequency domain. Instead, the Fast Fourier Transform (FFT) was chosen for the increase in preprocessing speed. Thus, the hyperparameter *window size* is chosen as $windowsize = 2^n$ with n is a positive integer and *window size* is bound by $windowsize \leq number\ of\ data$. After that, the data was padded with the size according to the *window size*.

Because each feature of a data was not completely independent, the ground velocity varies in different ranges [Min, Max] of different datasets, performing Min Max Scaling or Normalization were needed on all Fourier transformed features. After that, the Min and Max values was added as two new features for each data.

Next, the transformed data set were randomly divided into the Training set, the Cross-Validation set and the Test set with ratio 8:1:1. Then, the Standardization procedure were performed on each feature so that they have $mean = 0$ and $standard\ deviation = 1$. That was the final step of the preprocessing procedure.

After that, the model was trained with the preprocessed Training set using the Sklearn SVM library.

Finally, the hyperparameter *window size* was tested and evaluated on the classification model. Since hyperparameters depend on the characteristic of the dataset, the *window size* was chosen by testing with different value on the dataset to see which would bring the highest accuracy. To test the effect of the different

window sizes with different dataset size, a test on multiple simulated datasets were made and the result are shown in Sect. 4.

4 Result and Discussion

4.1 Result

Simulated Data Generation. After choosing the size of the data set as 2048 (2^{11}) number of data, 40 simulated datasets were generated with the randomized starting point, velocity, Non-linear moving tracks as presented in Sect. 3.1. A results of the moving track and simulated signals in the time domain were shown in Fig. 6.

A result of experiment data were shown in Fig. 7.

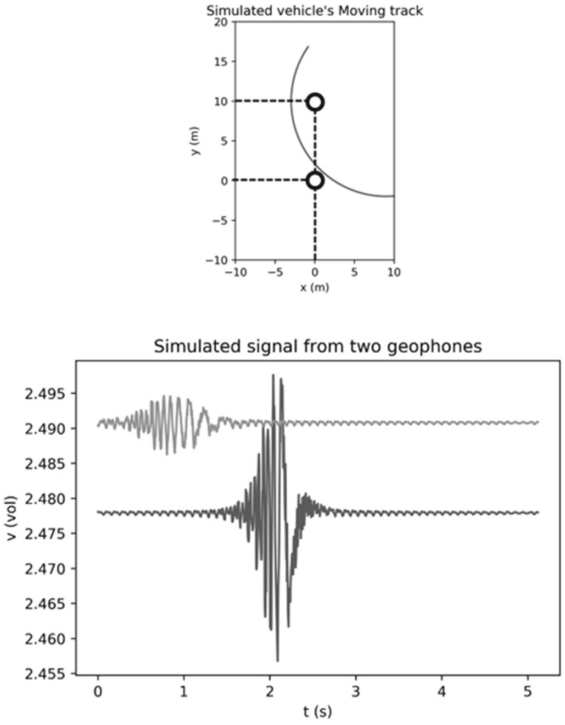

Fig. 6. A sample of track and simulated signal in the time domain.

Data Classification. The effect of the different *window sizes* with different number of SIMULATED dataset can be seen in Table 3:

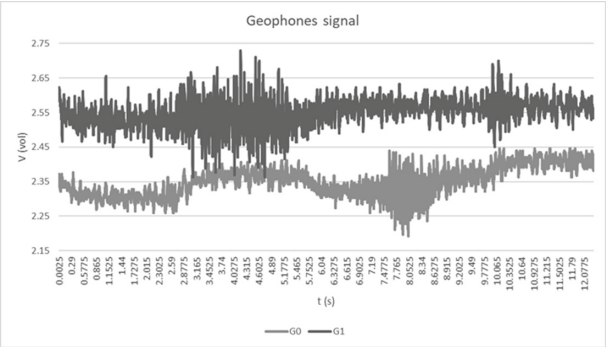

Fig. 7. A sample of track and simulated signal in the time domain.

Table 3. (SIMULATED DATA) Accuracy with different *window sizes* and different number of training datasets

		Window size (2^n)								
		1	2	3	4	5	6	7	8	9
No. of training dataset	5	87.60	93.46	96.00	97.75	98.83	**99.02**	96.88	93.36	88.28
	10	87.60	94.78	97.07	99.02	99.02	**99.17**	98.29	95.85	91.11
	15	87.01	93.42	97.20	98.70	**99.19**	99.12	98.57	96.58	92.84
	20	86.35	94.19	96.97	98.56	98.80	**99.39**	98.93	97.56	94.53
	25	86.99	94.39	97.01	98.81	**99.18**	99.02	**99.18**	97.38	95.02
	30	86.10	93.96	97.51	98.31	**99.19**	**99.19**	98.96	97.84	94.84
	35	85.81	93.71	97.15	98.63	**99.26**	**99.26**	99.20	97.81	94.53
	40	86.01	93.63	97.16	98.68	99.05	**99.28**	99.13	97.96	95.47

Table 4. (EXPERIMENT DATA) Accuracy with different window sizes and different number of training datasets

		Window size (2^n)								
		1	2	3	4	5	6	7	8	9
No. of training dataset	1	98.83	98.83	98.59	98.59	98.83	**99.53**	99.06	97.66	96.96
	2	98.71	98.61	98.71	97.64	**99.04**	98.93	97.86	93.35	93.89
	3	**97.30**	97.18	95.92	95.23	94.48	94.25	94.66	95.00	96.84
	4	93.83	94.68	94.60	94.43	95.02	94.09	94.85	96.38	**97.32**
	5	**95.61**	95.34	**95.61**	95.14	95.47	95.20	93.51	93.95	95.03
	6	**95.12**	93.97	94.59	94.64	94.28	94.19	92.51	90.63	92.15
	7	**94.96**	93.69	93.11	91.12	90.87	89.60	89.80	90.52	90.44
	8	**93.90**	93.81	92.11	90.07	89.35	88.42	87.02	87.67	88.24
	9	**94.90**	93.87	93.23	90.76	90.32	88.03	87.97	88.82	88.84

The effect of the different *window sizes* with different number of EXPERIMENT dataset can be seen in Table 4:

The predicting accuracy of the system trained with all nine EXPERIMENT datasets are shown in Fig. 8 in which the dash-line shown training accuracy while the other shown cross-validation accuracy.

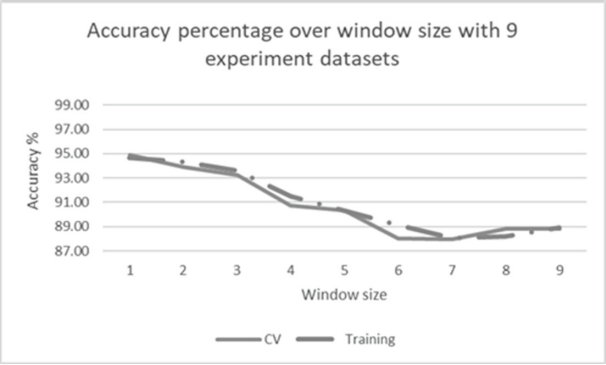

Fig. 8. Accuracy over *window size* with nine datasets.

Choosing the *window size* that brought the highest accuracy in a large size dataset, the *window size* $= 6$ was chosen for SIMULATED data and *window size* $= 1$ for EXPERIMENT data. To not be biased by the cross-validation data set, a final analysis on a Blind Test Set that the model has never seen before was used to evaluate the system and result in the accuracy of 99.10% for SIMULATED data and 94.22% for EXPERIMENT data.

Some conclusions can be made in Fig. 8 and Table 4. Overall, the predicting accuracy of the system are very high. Beside, there were a slight decrease in the accuracy when more datasets was added and the more datasets were used, the smaller the *window size* would bring the highest accuracy. This result implied that the current datasets were still not large enough to generalize the model to reach its best potential. Thus, acquiring more data and combining with the artificial datasets to train the model would much likely increase the predicting accuracy.

4.2 Discussion

At present state, a working system have been constructed to collect and analyze seismic signal generate from a moving vehicle. Preprocessing and Machine Learning techniques are used to classify acquired data thus determined whether the vehicle have passed the threshold or not.

The tests has been conducted only on a common road at HUST university using a 2015 KIA Morning car. Further tests are needed for integrating the artificial data and acquiring more data with different vehicle types to develop more robust analyzing system. With a small modification, the system can be made to work in real-time that satisfy the requirements of a long-lasting, unnoticeable, robust vehicle detection system with large monitoring range.

Acknowledgment. We would like to express our gratitude to the staff members in the Department of Precision Mechanical and Optical Engineering, Hanoi University of Science and Technology, Vietnam for supporting the research. We also would like to show our gratitude toward Mr. Anh Nguyen, Mr. Binh Nguyen and Mr. Phuong Le for their help in building the hardware for the experiment.

References

1. Aizerman, M.A.: Theoretical foundations of the potential function method in pattern recognition learning. Autom. Remote Control **25**, 821–837 (1964)
2. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 144–152. ACM (1992)
3. Choe, H.C., Karlsen, R.E., Gerhart, G.R., Meitzler, T.J.: Wavelet-based ground vehicle recognition using acoustic signals. In: Wavelet Applications III, vol. 2762, pp. 434–446. International Society for Optics and Photonics (1996)
4. Cortes, C., Vapnik, V.: Support vector networks. Mach. Learn. **20**, 273–297 (1995)
5. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the world with wireless sensor networks. In: Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP 2001), vol. 4, pp. 2033–2036. IEEE (2001)
6. Krylov, V.V.: Computation of ground vibrations generated by accelerating and braking road vehicles. Modal Anal. **2**(3), 299–321 (1996)
7. Krylov, V.V.: Generation of ground elastic waves by road vehicles. J. Comput. Acoust. **9**(03), 919–933 (2001)
8. Lan, J., Nahavandi, S., Lan, T., Yin, Y.: Recognition of moving ground targets by measuring and processing seismic signal. Measurement **37**(2), 189–199 (2005)
9. Sabatier, J.M., Xiang, N.: An investigation of acoustic-to-seismic coupling to detect buried antitank landmines. IEEE Trans. Geosci. Remote Sens. **39**(6), 1146–1154 (2001)
10. Stotts, L.B.: Unattended-ground-sensor-related technologies: an army perspective. In: Unattended Ground Sensor Technologies and Applications II, vol. 4040, pp. 2–11. International Society for Optics and Photonics (2000)