Applying Wavelet Packet Decomposition and One-Class Support Vector Machine on Vehicle Acceleration Traces for Road Anomaly Detection

Fengyu Cong[1], Hannu Hautakangas[1], Jukka Nieminen[1], Oleksiy Mazhelis[2], Mikko Perttunen[3], Jukka Riekki[3], and Tapani Ristaniemi[1]

[1] Department of Mathematical Information Technology, University of Jyväskylä, Finland
[2] Department of Computer Science and Information Systems, University of Jyväskylä
[3] Department of Computer Science and Engineering, University of Oulu, Finland
{Fengyu.Cong,Oleksiy.Mazhelis,Tapani.Ristaniemi}@jyu.fi,
{Hannu.Hautakangas,Nieminen.Jukka}@gmail.com,
{Mikko.Perttunen,JPR}@ee.oulu.fi

Abstract. Road condition monitoring through real-time intelligent systems has become more and more significant due to heavy road transportation. Road conditions can be roughly divided into normal and anomaly segments. The number of former should be much larger than the latter for a useable road. Based on the nature of road condition monitoring, anomaly detection is applied, especially for pothole detection in this study, using accelerometer data of a riding car. Accelerometer data were first labeled and segmented, after which features were extracted by wavelet packet decomposition. A classification model was built using one-class support vector machine. For the classifier, the data of some normal segments were used to train the classifier and the left normal segments and all potholes were for the testing stage. The results demonstrate that all 21 potholes were detected reliably in this study. With low computing cost, the proposed approach is promising for real-time application.

Keywords: Anomaly detection, one-class, pothole, road, support vector machine, wavelet packet decomposition.

1 Introduction

In recent years, road condition monitoring has become a popular research area due to intensive and still growing traffic that puts the road surface to constant stress. Intelligent systems for detecting bad road surface conditions can assist drivers to prevent damaging vehicles and even accidents and assist road management departments in timely discovering the need of maintenance on dangerous road conditions in time. Using GPS data and the acceleration measured by the accelerometer attached on some part of the driving car have been particularly used in the pothole and other anomalies detection for road management [9], [10], [17], [19]. Fig.1 shows the image of a pothole in the road [10] and Fig.2 demonstrates the accelerometer orientation [19]. This pioneering work has many advantages, especially in reducing the cost of the road management in

C. Guo, Z.-G. Hou, and Z. Zeng (Eds.): ISNN 2013, Part I, LNCS 7951, pp. 291–299, 2013.

contrast to many other systems including laser profilometer measurement [12], ground penetrating radar [14], collection and analysis of images of road segments [4], [11], [15], [18], and so on. Furthermore, with such work, real-time intelligent systems become possible.

Fig. 1. Image of pothole [10]

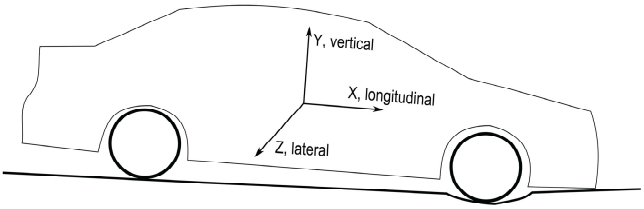

Fig. 2. Accelerometer orientation

In [10], [17] accelerometer data were used for classifying road segments as the normal or containing anomaly by applying simple threshold values and a high-pass filter. In [19], classification was realized by machine learning methods, namely, with multi-class support vector machine (SVM) [7] on time-domain and frequency-domain features. However, these studies did not fully consider the properties of anomaly in the road condition monitoring. In this study, we propose a novel approach involving the feature extraction, feature selection and the classification with the consideration of the nature of detecting potholes in the road.

For the feature extraction, two types of information tend to be used according to the mechanism of the data formulation and the transformation of the collected data. Particularly, in the time or frequency domain, the peak amplitude of the collected data within certain duration and the power spectrum density (PSD) are often measured [10]. The peak amplitude is very sensitive and may be affected by many factors since it just contains the information within a very short duration. The PSD actually assumes the data are stationary [16], and this assumption might be correct when the road is smooth and the collected accelerations tend to be stationary. However, in case that a car passes the anomaly, for example, the pothole, the collected accelerations are transient in short duration and they are definitely not stationary any more [10]. Thus, the transformation for the non-stationary signal should be used in the feature extraction for the anomaly. From this point of view, the wavelet transformation based methods are appropriate candidates for the feature extraction [8]. Hence, the wavelet packet decomposition (WPD) is used for the feature extraction in this study [21].

The feature selection plays an important role in pattern recognition [1]. It assists to remove the redundant features and obtain the discriminative features. We try four of the generally used methods in this study. They are forward selection (FS) [1], backward selection (BS) [1], genetic algorithm (GA) [22] and principal component analysis (PCA) [1]. After comparing the performance in detecting potholes in the road, we may determine the best method for feature selection in our study.

After the feature is extracted, a machine learning based classifier is usually exploited to recognize the feature. There are mainly three ways to construct such

classifiers including the supervised, the semi-supervised and the unsupervised. For an anomaly detection problem, there are often two classes of samples to recognize [2]. However, since the number of samples of the anomaly is usually much smaller than that of the normal, the supervised method is not appropriate, but the semi-supervised and the unsupervised methods are [2]. The semi-supervised approach implies that the classifier is trained with normal data, and then tested with an independent set of normal and anomaly data. This is often named as the one-class classification which is also referred as outlier detection, novelty detection or anomaly detection [2]. In this problem, the data instances that do not belong to the normal data are separated. Here, the one-class SVM [20] is applied.

In this study, accelerometer data were collected by the system used in [19]. Part of one car's normal data was used for training the classifier and the left normal data of that car and pothole data of three cars were for the testing stage. All 21 potholes were successfully detected by the proposed approach.

2 Method

2.1 Data Description

In order to make reliable analysis one must ensure high quality of the data. Further data analysis is much harder or even impossible if the data is invalid or badly corrupted. The accelerometer data in this work were collected by Perttunen et al. from the University of Oulu in Finland. The data were collected using a Nokia N95 mobile phone mounted on the car's dashboard. The N95 mobile phone has a built-in 3-axis accelerometer and a GPS receiver. The sampling frequency of the accelerometer was 38 Hz. The GPS data were not used in this study.

Three different cars were used to collect data. Each drive was also recorded using a video camera, which was attached to the head rest of the passenger's seat. The video was synchronized with accelerometer data after the drive so that it could be seen on the video when the car hits a road anomaly, for example a speed bump. This allows marking certain measurements of the accelerometer as anomalies. During the video analysis, it was noticed that it was impossible to observe when exactly the anomaly begins and ends in the level of milliseconds. It was also hard to classify the anomalies into different categories at the same time. Of course, some of the categories are easy to classify, for example speed bumps associated with potholes. But other kinds of anomaly were not easy to recognize from the video. So, the classification in this study was not targeted to recognize different potholes, but to discriminate potholes from normal road conditions, i.e., to detect potholes.

While the data were labeled, 21 segments with potholes were extracted from three drives for investigation, and 1764 normal segments from one car were used for training and testing the classifier. Normal segments of the other two roads were not used for analysis since the roads were not very smooth. The duration of each segment was three seconds and a Hamming window was used. Each segment of the pothole was arranged in the center of the segment to avoid energy leakage. Although accelerations in three directions were measured, we found the y-axis data mostly revealed the changes of

accelerations when a car passed a pothole in contrast to the normal road conditions. Therefore, accelerations in this direction were chosen for analysis in this study. Fig.3 shows two segments containing the maximum and minimum powerful potholes and normal segments. The data were filtered by a 1-5 Hz band-pass filter for better visualization.

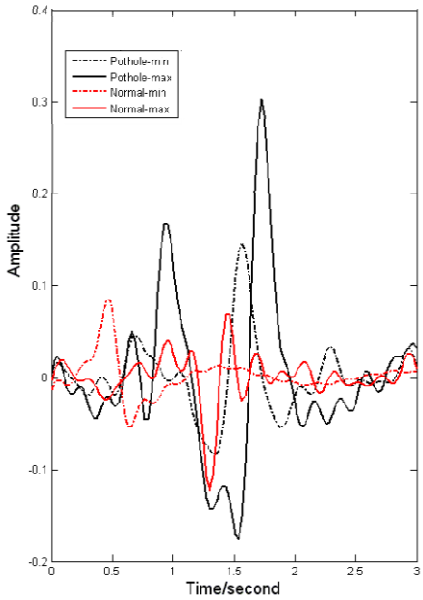

Fig. 3. Accelerations of potholes and normal segments

2.2 Feature Extraction

Feature extraction is by WPD here. The decomposition of a signal by WPD can be illustrated by successive low-pass and high-pass filters shown in Fig.4 since WPD can be carried out by an iterated application of quadrature mirror filters and followed by downsampling [21]. The variance of coefficients at each end node can be regarded as a feature of the signal [21]. For this approach, the key questions for WPD are how to determine the type of wavelets and the number of levels for decomposition and how to select coefficients at the desired end notes for formulating features [5].

Feature extraction by WPD is indeed based on the wavelet transform, which can also be regarded as a filtering process. Therefore, the frequency responses of such filter should match the spectral properties of desired signals [5], [6]. In this study, we used Daubecheis wavelet [8] since the frequency responses of this wavelet under the selected parameters for WPD match the spectral properties of accelerations of potholes. Indeed, this wavelet is very generally used for wavelet based analysis. The number of samples in each segment is 114 as the sampling frequency is 38 Hz and the duration of window is 3 seconds. For WPD, two to power of the number of levels should approximate the number of samples [5]. Therefore, seven levels were determined in this study. And then, 128 features were exacted and distributed from 0-38 Hz (Note: sampling frequency is 38 Hz here). Furthermore, after checking the power spectrum of potholes and normal segments, we decided to select the WPD features within the frequency range 1.5-10 Hz. Subsequently, the 30 features from # 5 to # 34 among 128 features [5] were finally chosen for the further pothole detection. Then, each feature was normalized to its

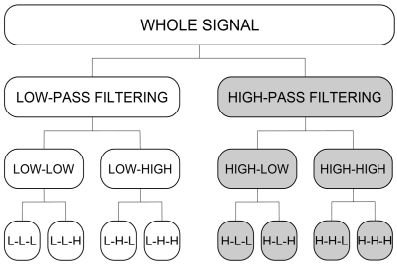

Fig. 4. Multi-scale decomposition for WPD [21]

standard deviation before feature selection. Toolbox of Wavelet in MATLAB (The Mathworks, Inc., Natick, MA) was used for WPD.

2.3 Feature Selection

Although some features were selected among all extracted features during the feature extraction process in terms of the rough evaluation of power spectrum of potholes and normal segments, there were still 30 features left for classification. Indeed, such a number is still too high for real-time application. Therefore, machine learning based feature selection can assist to reduce the computing demand for classification. Feature selection was designated to find the features best discriminating the potholes and normal segments in this study. We tried four extensively used methods here. For the completeness of the study, they are briefly introduced next. Please refer to [1] for more details.

Forward selection [1] is one of the basic feature selection methods. As a part of stepwise selection method, the idea, like in other feature selection methods, is to select a subset of features, which yields accurate enough results compared to results with all features. Forward selection begins with zero features in the model. The first feature is selected by testing each feature and then selecting the feature that yields, for example, the best classification result or the best f-value in statistical tests, such as, ANOVA (analysis of variance). This feature is the most significant feature. When a feature is selected, it is moved from unselected feature set to selected feature set. After this, the algorithm continues by comparing which feature of the remaining unselected features yields the best result with previously selected features and moves that feature from the unselected to the selected feature set. The procedure ends when there are no more features that increases the result or increases it only a bit. Another ending condition can be by a pre-defined upper number of selected features.

Another popular feature selection method is backward selection [1], which is opposite process compared to the forward selection. While forward selection begins with zero selected features, backward selection begins with all features. In every round, the algorithm tests all of the remaining features and removes the feature that decreases the results the most. This procedure continues until the result does not increase or it increases only a bit or the pre-defined number of features reaches.

The model of genetic algorithms was introduced by John Holland in 1975 [13]. Genetic algorithms are a group of computational models searching a potential solution to a specific problem using a simple chromosome-like data structure, which is inspired by evolution [1], [22]. A chromosome is a set of instructions which one algorithm will use to construct a new model or a function, such as, an optimization problem or selecting a subset of features for SVM. All features are represented as a binary vector of size m, where m is the number of features. '1' means that a feature is part of the subset and '0' means that the feature is not part of the subset. An algorithm can be considered as a two-stage process. It begins with the current population where the best chromosomes are selected to create an intermediate population. Recombination and mutation is then applied to create the next population. This two-stage process constitutes one generation in the execution of a genetic

algorithm. The algorithm begins with initial population of chromosomes. Typically initial population is chosen randomly from the original dataset. Then each chromosome is evaluated and assigned a fitness value. Chromosomes which represent better solution for the target problem are given a better fitness value than those chromosomes that provides poorer solution. Better fitness value means better reproducing chances. Reproducing may occur through crossover, mutation or reproduction operations. Please refer to [22] for more details about using an genetic algorithm for feature selection.

PCA is another frequently used method for feature selection [1]. The object of PCA is to find uncorrelated principal components that describe the dependencies between multiple variables. The principal components are ordered so that the first component explains the largest amount of variance in the data, and the second component is for the second largest variance, and so on. Generally it is expected that most of the variance in the original data set is covered by the first several principal components. PCA can be easily produced through the eigenvalue decomposition of the covariance matrix of multivariate datasets. It must be noticed that forward and backward selection and genetic algorithm do not affect data, and they are just methods to choose the best feature combinations, but PCA is for new features.

2.4 One-Class SVM

SVM is probably the mostly used classifier in the past decade [3]. It is a binary classifier that was invented by Cortes and Vapnik in 1995 [7]. It is based on generalized portrait algorithm and on Vapnik's research in statistical learning theory from the late 1970's. Basically SVM is intended to classify only two classes but it can be extended to support one-class and multiclass classification.

One scheme of one-class SVM is to map the input data into a high dimensional feature space and then fit all or most of the data into a hypersphere [3]. The volume of the hypersphere is minimized and all of the data samples that do not fall in the hypersphere are considered as anomalies. Another idea is the ν-SVM that creates a decision function that separates most of the training data from the origin with a maximum margin [3]. The parameter ν is associated with the number of support vectors and outliers. So, one-class SVM tries to find a separating hyperplane and maximizes the distance between the two classes while two-class SVM can be solved by constructing a hypersphere that captures most of the training data and minimizes its volume or separates training data from the origin with maximum margin [3].

2.5 Data Processing

LIBSVM software was used [3]. We used one-class SVM with Gaussian radial basis kernel with the parameters ν and γ with values 0.01 and 0.00002, correspondingly. SVM was trained with 70% of the normal data (1234 segments) and the training data was chosen randomly. Rest of the normal data (530 segments) and all 21 anomaly segments were used to test the accuracy of the constructed SVM model.

We present the result using sensitivity and specificity derived from the confusion matrix [1]. Such a matrix includes true positive (TP: 'pothole' is classified as 'pothole'), false negative (FN: 'pothole' is classified as 'normal'), false positive (FP: 'normal' is classified as 'pothole') and true negative (TN: 'normal' is classified as 'normal'). Sensitivity is equal to TP/(TP+FN), and specificity is defined as TN/(TN+FP). Their ideal values are 1 with zero FN and zero FP.

3 Results

All classification results that we present are the average of 1000 SVM classifications. If all 30 features were used for the classification, TP was 21, i.e., all 21 potholes were recognized as potholes, and TN was about 524 and FP was about 6, i.e., six normal segments were recognized as potholes.

For feature selection, we tested the four methods mentioned in the subsection 2.3 using different numbers of selected features. The results are shown in Fig.5. Obviously, PCA is the best for feature selection when the number of features is larger than 5. FS is the best when the number of features is larger than 2 and smaller than 6.

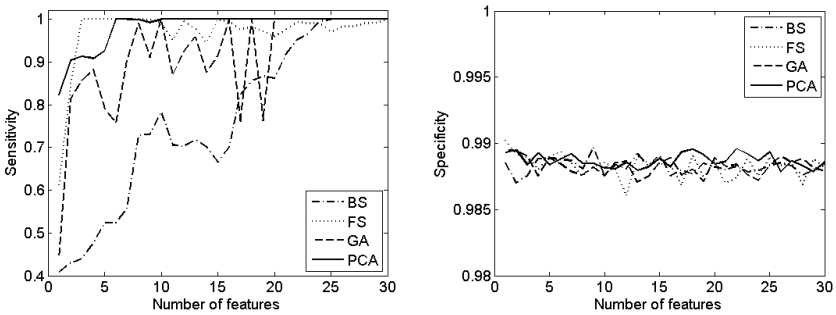

Fig. 5. (a) Sensitivity (b) specificity

4 Conclusion

This study has attempted to solve road surface anomaly detection problems with mobile phone's embedded accelerometer, wavelet packet decomposition for feature extraction, feature selection methods, and one-class support vector machine for classification. The achieved results are promising. As shown in Fig.5, one-class anomaly detection could be done very accurately. Our best true positive rate was 100%. These results are surprisingly good with our limited and unbalanced data sets.

In the data collection, some of the timestamps were corrupted and anomalies mislabeled due to limitation of hardware. This gives some extra challenge to data analysis. Thus, development of a proper data collection framework is essential in

order to obtain reliable and stable results for real-time classification system in real production environment.

Acknowledgments. The work was supported by Sensor Data Fusion and Applications in Cooperative Traffic ICT SHOK project (TEKES, Finnish Funding Agency for Technology and Innovation).

References

1. Alpaydin, E.: Introduction to machine learning, 2nd edn. MIT Press (2010)
2. Chandola, V., Banerjee, A., Kumar, A.: Anomaly detection–A Survey. ACM Computing Surveys 41 (2009)
3. Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology 2, 1–27 (2011)
4. Cheng, H.D., Miyojim, M.: Automatic Pavement Distress Detection System. Information Sciences 108, 219–240 (1998)
5. Cong, F., Huang, Y., Kalyakin, I., et al.: Frequency Response Based Wavelet Decomposition to Extract Children's Mismatch Negativity Elicited by Uninterrupted Sound. J. Med. Biol. Eng. 32, 205–214 (2012)
6. Cong, F., Leppanen, P.H., Astikainen, P., et al.: Dimension Reduction: Additional Benefit of an Optimal Filter for Independent Component Analysis to Extract Event-Related Potentials. J. Neurosci. Methods 201, 269–280 (2011)
7. Cortes, C., Vapnik, V.: Support-Vector Networks. Mach. Learn. 20, 273–297 (1995)
8. Daubechies, I.: Ten lectures on wavelets. Society for Industrial and Applied Mathematics (1992)
9. DuPont, E.M., Moore, C.A., Collins, E.G., et al.: Frequency Response Method for Terrain Classification in Autonomous Ground Vehicles. Auton. Robot. 24, 337–347 (2008)
10. Eriksson, J., Girod, L., Hull, B., et al.: The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. In: Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services (MobiSys 2008), Breckenridge, CO, USA, pp. 29–39 (2008)
11. Georgopoulos, A., Loizos, A., Flouda, A.: Digital Image Processing as a Tool for Pavement Distress Evaluation. ISPRS Journal of Photogrammetry and Remote Sensing 50, 23–33 (1995)
12. Hesami, R., McManus, K.J.: Signal Processing Approach to Road Roughness Analysis and Measurement, pp. 1–6 (2009)
13. Holland, J.: Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor (1975)
14. Houston, D.R., Pelczarski, N.V., Esser, B., et al.: Damage Detection in Roadways with Ground Penetrating Radar 4084, 91–94 (2000)
15. Lin, J., Liu, Y.: Potholes Detection Based on SVM in the Pavement Distress Image, pp. 544–547 (2010)
16. Mitra, S.: Digital signal Processing—A computer based approach. McGraw–Hill (2005)
17. Mohan, P., Padmanabhan, V.N., Nericell, R.: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones, pp. 323–336 (2008)
18. Nejad, F.M., Zakeri, H.: A Comparison of Multi-Resolution Methods for Detection and Isolation of Pavement Distress. Expert Syst. Appl. 38, 2857–2872 (2011)

19. Perttunen, M., Mazhelis, O., Cong, F., et al.: Distributed Road Surface Condition Monitoring using Mobile Phones. In: Hsu, C.-H., Yang, L.T., Ma, J., Zhu, C. (eds.) UIC 2011. LNCS, vol. 6905, pp. 64–78. Springer, Heidelberg (2011)
20. Rajasegarar, S., Leckie, C., Bezdek, J.C., et al.: Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks. IEEE Trans. Inf. Forensic Secur. 5, 518–533 (2010)
21. Schclar, A., Averbuch, A., Rabin, N., et al.: A Diffusion Framework for Detection of Moving Vehicles. Digit. Signal Prog. 20, 111–122 (2010)
22. Yang, J., Honavar, V.: Feature Subset Selection using a Genetic Algorithm. IEEE Intell. Syst. 13, 44–49 (1998)