Probabilistic classifiers and time-scale representations: application to the monitoring of a tramway guiding system

Zahra HAMOU MAMAR[1], Pierre CHAINAIS[1] and Alexandre AUSSEM[2]

1- LIMOS, University Clermont-Ferrand II - CNRS UMR 6158
ISIMA, BP 10125, Campus des Cézeaux 63173 Aubière cedex - France
hamou@isima.fr, pchainai@isima.fr

2- COMAD Team, PRISMa Lab., University Lyon 1,
8, boulevard Niels Bohr 69622 Villeurbanne cedex - France
aaussem@univ-lyon1.fr

Abstract. We discuss a new diagnosis system combining wavelet analysis techniques and probabilistic classifiers for detecting tramway rollers defects. A continuous wavelet transform is applied on the vibration signals measured by specific accelerometers located on the rail. A temporal segmentation of the signals is carried out in order to identify the contribution of each pair of rollers to the overall vibration signal. The singular values decomposition (SVD) method is applied to segments of the time-scale representation to extract the most significative features. The resulting multi-class problem is then solved using pairwise classifiers trained on two-class sub-problems. The efficiency of this approach is successfully illustrated on several experiments on the tramway.

1 Introduction

Most often, the monitoring of mechanical parts of a transportation system is based on visual controls performed by experts or technicians. Such tasks are expensive and time consuming ; moreover the diagnosis is sometimes unreliable. The use of an automatic monitoring system reduces the maintenance costs and makes them more efficient. Several works deal with the diagnosis, detection and classification of defects of a rail [9, 10, 11] or of mechanical systems like roller bearings [7, 8]. This paper focuses on the monitoring of the rollers of the guiding system of a new kind of a tramway on tires, the Translohr.

The Translohr, fig. 1(a), has been chosen by several cities (Clermont-Ferrand, Venice, Padova...). It is equipped with a guiding system consisting of a single central rail. The tramway used for our experimental tests has 4 axles equipped on both sides with two pairs of rollers at the front and at the back, fig. 1(b)&(c). There are therefore 8 pairs of rollers. To limit noise and parasitic vibrations, the rollers are covered with a composite overlay that ensures better comfort and silence without squeaks. This overlay can be damaged and worn, fig. 1(d). The guiding system then becomes quite uncomfortable and the structure of the rollers themselves is threatened: a renewal of the overlay must be decided.

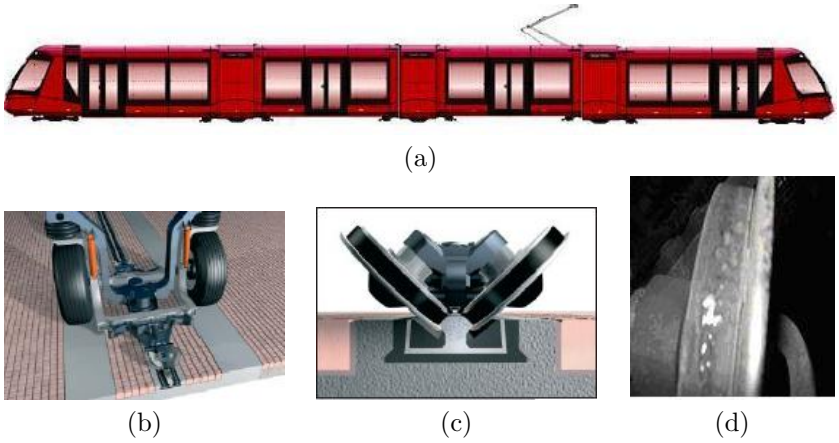


Fig. 1: *(a)* The Translohr on tires. *(b)* Overview of an axle and of the guiding system. *(c)* The guiding system: a pair of rollers. *(d)* A defected pair of rollers.

In the framework of a project developed and supported by the city of Clermont-Ferrand (SMTC), LOHR and Signal Development companies, we develop a system designed for the automatic monitoring of the state of the overlay of the rollers. This system uses the measurement of the vibrations[1] provoked by a damaged overlay in the rail. Thanks to a time-scale analysis of the signals combined to learning methods, one hopes to locate the defective pair of rollers and to identify the nature of the defects at the same time. The state of each pair of rollers must be determined as precisely as possible: new (no defect), (slightly or very) worn, with holes (due to the accidental presence of metallic pieces or stones on the rail)...This last classification task can be performed by using several classifiers types. We have compared results from MLP, Knn and RBF. Three states have been discriminated in the results presented below: "new", "worn" and "with holes", see Section 4.

2 Signal processing

2.1 Data pre-processing

We have recorded several sets of signals for different known configurations (combinations of "new", "worn" and "with holes" rollers on different axles). The horizontal and vertical vibrations of the rail are measured by three bidirectional accelerometers when a Translohr passes at normal speed (35 Km/h). The use of three sensors ensures some redundancy. The measurement system provides us with six signals (3 horizontal and 3 vertical components). Each acquisition lasts 8 seconds at sampling frequency $Fe = 44100Hz$. Thus, the available bandwidth for analysis is roughly within [2Hz-20kHz]. In fact, only the sum of the two

[1]The measurement system is the propriety of Signal Development SD company.

components (lateral and vertical) is used.

The signal is a priori related to information coming from the 8 pairs of rollers that generate vibrations altogether, at the same time. However, it is a reasonable physical assumption to consider that only the closest pair of rollers to the sensor generates the measured vibrations. Two photoelectric cells around the sensors provide references in time and position. This information allows us to locate the passage of each of the 8 pairs of rollers over the sensor within each signal. Therefore, 8 segments of signal can be identified and labelled CG_k, for $k = 1, ..., 8$, fig. 2(a).

2.2 Time-scale representation

Thanks to the procedure described in the previous section, we are able to know to which pair of rollers is related each segment of the signal. Now, we need to identify the state of the overlay of this pair of rollers. To this aim, a local frequency information is required. Such a remark naturally leads to use a time-frequency or a time-scale analysis.

Two major types of defects can be distinguished. The first type corresponds to structural roller defects that generate vibrations at low frequencies (or large scales equivalently). The second one generates high frequencies (small scales) and is characterized by surface defects, fig. 1(d). To get a good frequency resolution both at low and high frequencies, a time-scale analysis is preferable: the relative precision is constant over the whole range of analyzed frequencies ($\Delta\nu/\nu$ is constant). We used a continuous wavelet transform of the form [2]:

$$T_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-b}{a})dt \quad (1)$$

where $\psi(t)$ is the wavelet function, a is the scale or dilation parameter (scale $\approx$ 1/frequency), and b is the time location parameter. A scalogram such as presented on fig. 2(c) consists of a gray level picture of the energy density function of the wavelet transform, $|T_x(a,b)|^2$. The wavelet used in this work is the second derivative of the Gaussian function, known as the Mexican-hat wavelet [2].

Figure 2 shows a typical example of the scalogram of a signal. Some information are directly visible by eye. Let F_{rot} the rotation frequency of the rollers. The scalogram is clearly divided into three parts. At the bottom is a low frequency band, $f < F_{rot}$, corresponding to the first type of defects (warped roller, structural deformation...) ; the rotational frequency F_{rot} is represented by an horizontal dashed line at scale $a = 2^{12}$ corresponding to $10.76Hz$ (depending on the tramway speed). At the top is a high frequency band $f \gtrsim F_{noise}$ with $F_{noise} \simeq 350$ Hz (corresponding to the scale $a \simeq 2^7$), essentially consisting of noise. The medium frequency band corresponds to $F_{rot} \lesssim f \lesssim F_{noise}$, $f \in [10Hz - 350Hz]$. This intermediate band is the most informative. One clearly notices the existence of black peaks, fig. 2(c), look at zone A. Such peaks can be associated to defects of a size of the order of several millimeters to 10 centimeters on the overlay of a pair of rollers.

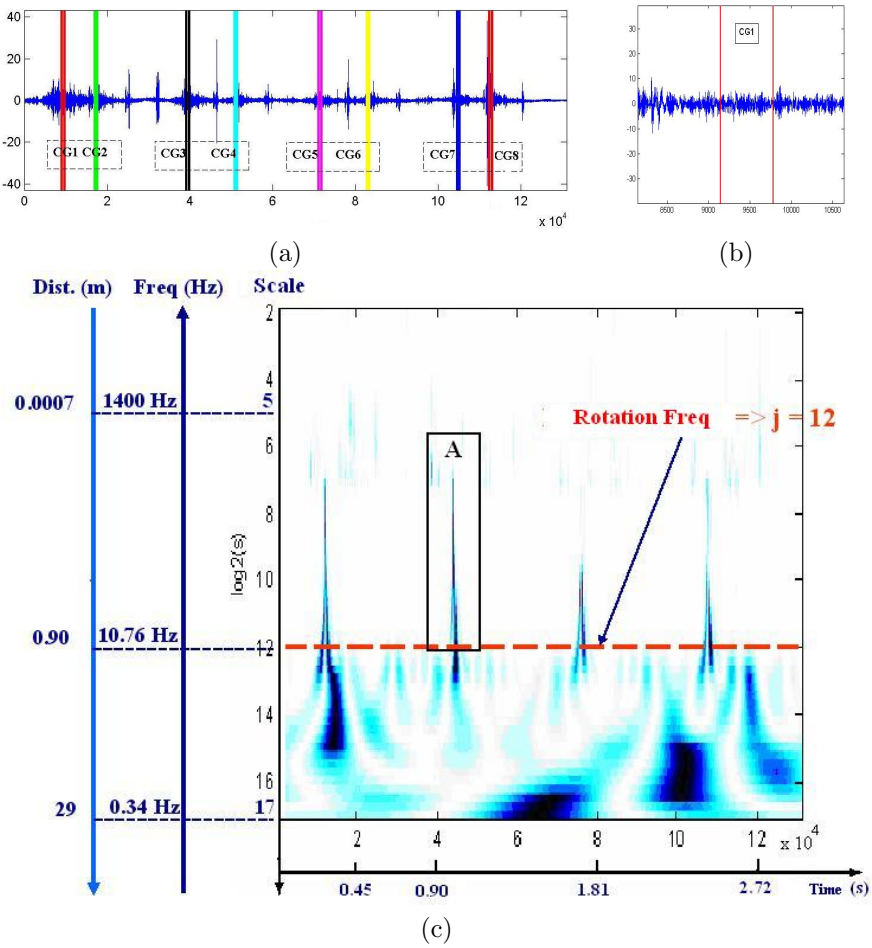


Fig. 2: *(a)* Whole temporal signal. *(b)* Zoom on the segment corresponding to pair of rollers CG1. *(c)* The time-scale representation (scalogram) of a signal.

3 Feature selection & Classification

In this first study, we will only focus on medium scale defects associated to the medium frequency band defined above. The time segmentation described in section 2.1 combined to the frequency segmentation described in previous section allow us to isolate a rectangular piece of scalogram for each pair of rollers. Such a piece of scalogram is indeed a matrix of coefficients. Taking all the wavelet coefficients for classification purpose is clearly impractical. A reduction of the dimensionality to significantly increase the generalization performance is necessary. Geometrical features of the observed dark peaks in fig. 2(c) such as pixel densities, surface of the peaks... are not sufficiently informative to get an efficient classification. A singular values decomposition (SVD) [3] of the scalogram yields more appropriate features. Singular values decomposition of time-frequency Wigner distributions was used for classification in [5, 6]. Let CWT_k the matrix of squared wavelet coefficients associated to CGk. Every CWT_k matrix can be decomposed as:

$$CWT_k \quad = \quad USV^t. \tag{2}$$

where U and V are orthogonal matrices and $S = diag(\lambda_1, \lambda_2, ..., \lambda_p), \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$.

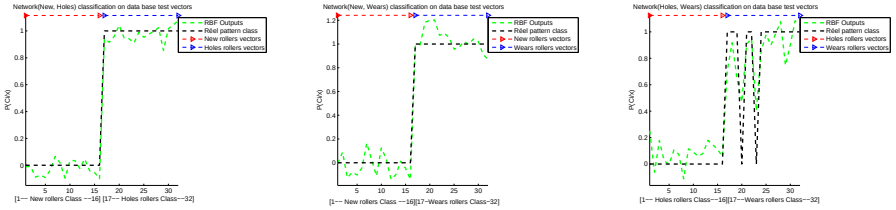
Due to the limited number of training data, the 3-class ("new", "worn", "with holes") classification problem has been decomposed into a set of 3 simpler 2-class problems following the approach described in [4]. Each network is trained using the data of two classes only to obtain a posterior probability for the class decision ("new" vs "worn", "new" vs "with holes", "worn" vs "with holes"). The three resulting probabilistic pairwise classifiers P_{ij} are combined afterwards in order to obtain posterior probabilities $P(C_i|X = x)$ for the final class decision:

$$P(\mathcal{C}_i|X = x) = \frac{1}{\sum_{j=1, j\neq i}^{K} \frac{1}{P_{ij}} - (K-2)} \tag{3}$$

where $K = 3$ in the present case and P_{ij} is the probability that x is assigned to class $\mathcal{C}_i$ given that it belongs to $\mathcal{C}_i \cup \mathcal{C}_j$.

4 Results of defects classification

80% of the vectors of our data basis were used for training. The 20% remaining vectors were kept aside for validation purpose. Several classification techniques were compared : k-nearest neighbours (Knn), neural network classifiers (MLP), and radial basis function networks (RBF). The two-class classification using RBF gave the best results. The application to our test bases gave very convincing results. The data for test consists of 3 sets of 16 vectors each ("new", "worn", "with holes"). Figure 3 shows the results of the two-class network classifiers. The outputs are plotted as a function of the index of vectors: the first 16 belong to one class and the other 16 to another one. The distinctions "new" *vs* "worn"



(a) (b) (c)

Fig. 3: RBF networks outputs : *(a)* new *vs* holes ; *(b)* new *vs* worn. *(c)* holes *vs* worn

and "new" *vs* "with holes" are perfectly efficient. The "worn" *vs* "with holes" classification is quite efficient as well with only 2 outliers over 32 tested vectors. Finally, the combination of the three pairwise probabilistic classifiers using eq. 3 leads to the following recognition rates: 100% for "new", 76% for "worn", and 83% for "with holes". We expect an increase of these performances thanks to a more important data basis and the use of SVM which is the subject of ongoing work. We also study the extension of this method to a larger number of classes with different levels of wear.

References

[1] C.M. Bishop. *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

[2] S. Mallat. *A Wavelet Tour of Signal Processing*, Academic Press, 1997.

[3] G.H. Golub, C.F. Van Loan. *Matrix Computations*, the John Hopkins University Press, Baltimore, London, 1989.

[4] D. Price, S. Kneer, L. Personnaz, G. Dreyfus, Pairwise neural network classifiers with probabilistic outputs, *Neural Information Proceedings Systems*, 1994.

[5] N.M. Marinovic, G. Eichmann, Feature extraction and pattern classification in space–spatial frequency domain, *SPIE Intell. Robots Comput. Vision 579, pages 19-25*, 1985.

[6] N.M. Marinovic, G. Eichmann, An expansion of Wigner distribution and its applications, *Proceedings of the IEEE ICASSP-85, Vol. 3, pages 1021-1024*, 1985.

[7] C. Junsheng, Y. Dejie, Y. YuA, fault diagnosis approach for roller bearings based on EMD method and AR model, *Mechanical Systems and Signal Processing, Volume 20, Issue 2, Pages 350-362*, February 2006.

[8] J.K. Hammond, P.R. White,The analysis of non-stationary signals using time-frequency methods, *Journal of Sound and Vibration, Volume 190, Issue 3, 29, Pages 419-447*, February 1996.

[9] M. Bentoumi, Diagnostic de systèmes non linéaires par réseaux de neurones - Application à la détection temps-réel de défauts de rail. Thèse de doctorat, Université Henri Poincaré de Nancy,France, 2004.

[10] L. Oukhellou, P. Aknin, Hybrid training of radial basis function networks in a partitionning context of classification. Revue Neurocomputing, Vol. 28, Nos. 1-3 pages 165-175, 1999.

[11] M. Bentoumi, G. Millérioux, G. Bloch, L. Oukhellou, P. Aknin, Classification de Défauts de Rail par SVM. Congrès International IEEE Signaux, Circuits et Systemes SCS'04, Monastir, Tunisie, 18-21 mars 2004.