Use of Combined Railway Inspection Data Sources for Characterization of Rolling Contact Fatigue

Shulin Gao, Thorsten Szugs and Ralf Ahlbrink
EURAILSCOUT Inspection & Analysis, Germany, ralf.ahlbrink@eurailscout.com

Abstract

Rolling contact fatigue is caused by interaction of train wheels with railheads. To minimize operational risks railways infrastructure owners can ask for regular inspections to determine critical sections of assets in a predictive maintenance scheme.

Some defect types, for instance (sub-)surface anomalies like squats, were herein assessed with the EURAILSCOUT train UST02 by specific ultrasonic (UT) and eddy current (ET) testing probes, but each with inherent limitations. Additionally surface imaging video (VT) was recorded.

Machine learning techniques were applied to combine data from these three inspection systems from a recorded track, which had a squat defect population. The target was to provide optimized classification as basis for prediction of rolling contact fatigue growth. Results are described of a 'combined systems method', CSM, with classification features derived from densely spaced clusters with input from UT B-scans, and parameters from ET and VT raw data after signal and image processing, respectively. In comparison to classification per single system, CSM gave significantly improved hit rate without degradation of specificity.

1. Introduction

Modern rail production methods and track building techniques caused a change in the distribution of rail defects found by traditional rail inspection methods. Nowadays, most defects are caused by higher speeds, axle loads and train frequencies. Such defects are situated, where the loads and forces are transferred to the rail, i.e. in the railhead and the running band of the rail.

Available inspection methods have different strengths and weaknesses to detect various anomalies of railheads (1). Combination of ultrasonic testing (UT) mainly for defects in bulk, and eddy current testing (ET) for a specific rolling contact fatigue (RCF) type, i.e. head checking (HC), is state of the art for about a decade (2). Visual inspection (VT) of railway tracks is also common part of maintenance strategies. For detection of RCF by video few recording systems are on the market. Automatic detection of squat flaws in video images is a topic of on-going research, see e.g. (3), (4) and (5).

We used earlier (6) a basic approach using clustering to combine results of inspection methods for detection of squat-like RCF defects. Based on the same UT, ET and VT data set, here an improved and more general 'combined systems method' (CSM) is presented.

2. Methodology

Figure 1 shows a generic scheme for combination of inspection data. Each method output needs to be matched properly to inspected track positions. The EURAILSCOUT train UST02 is equipped with UT, ET and VT systems, which are synchronized with a central localization device. The latter system in turn records calibrated odometer, receiver data of global navigational satellite systems (GNSS), and sensed turnout parts as input for an off-line routing process, which uses additionally railway track map data as reference.
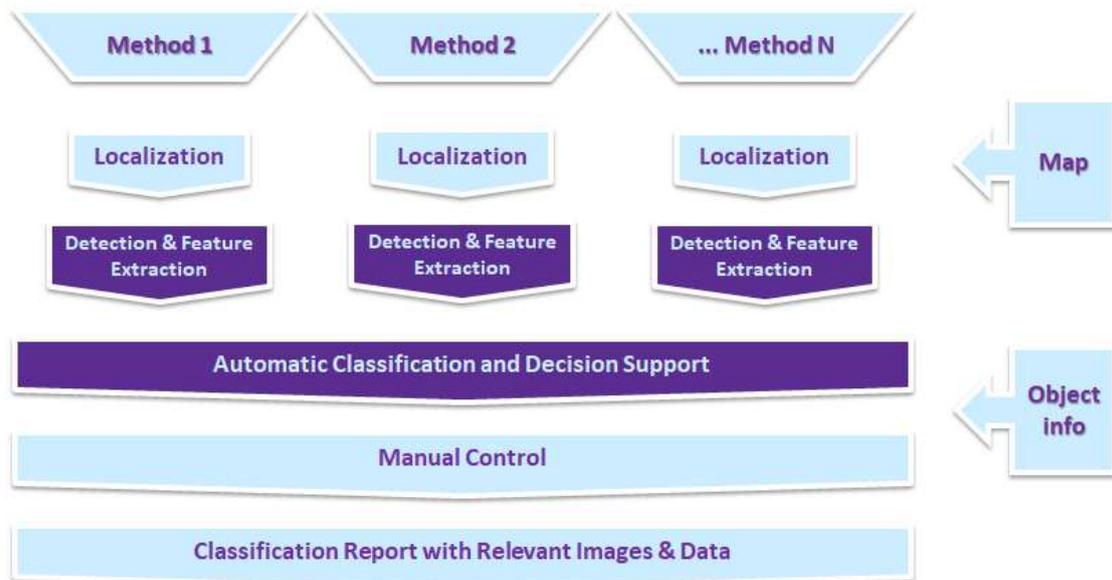

Figure 1: Scheme gives a generic overview for a combined evaluation approach. The developed prototype of 'combined systems method', CSM, is an implementation of the marked central steps.

Up to this stage, the described steps are part of an operational process. Subsequent steps traditionally comprise of viewer or analysis tools for individual method. This is done either with analysis performed by an operator, or for specific objects (e.g. welds), or for some rail defect types with automated sizing and classification based on data from a single system (e.g. head checks in ET).

Feature extraction per method and a joining automatic classification step are shown with dark background in Figure 1. These steps were implemented and evaluated in terms of classification performance.

The prototype application was programmed with usage of Python data science stack (7) and OpenCV (8). In particular, for the classification step a linear support vector machine with stochastic gradient descent learning (9) was applied.

3. Results

The data recording setup was the same as described in (6). The UT device is a 32-channel system, which sampled in tests with 2 mm spacing. The VT system records the top of the rail with a slight tilt to the gauge corner. For ET in particular, a non-standard probe arrangement was used, so that the probe coverage started from the middle of the rail going inwards to the gauge side, covering about 24 mm of the rail contour.

To assess the classification performance of the implemented method two data sets of two inspection runs on same track sections were used. 30 squat mid positions were determined by manual tagging from video data as reference input. Partially we left out small suspects, which might have been 'small' squats, i.e. at early stage, because the defects were not verified on-site in the past. This is a shortcoming of the currently used evaluation, which potentially cause a slight overestimation of false positive rate; and furthermore it is the reason why this investigation gives no proper insight on performance for small squats.
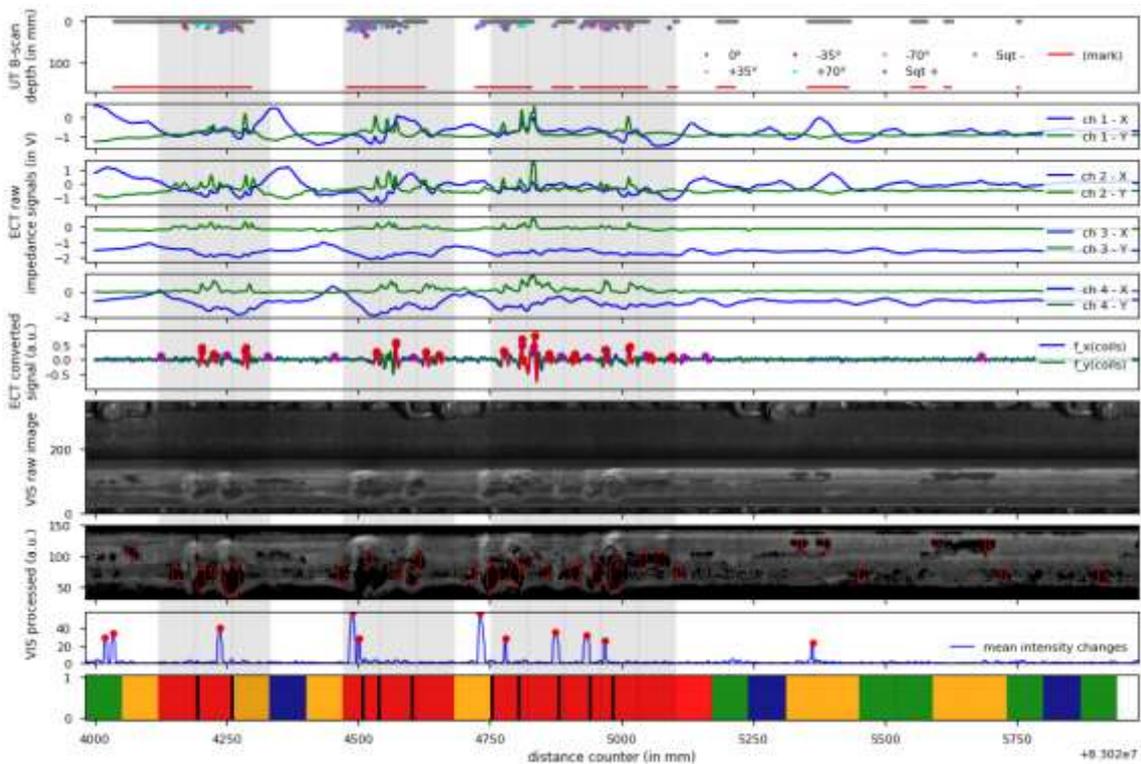

Figure 2: Top lane (#1) – UT B-scan, lane #2-#6 – ET raw XY and converted signals, lane #7-#9 – VT raw and processed image with Blob detection, plus mean intensity changes; and bottom lane (#10) – clustering with colour code indicate hits per system, and black lines are squat reference positions. Red marks are indicators for diverse feature parameters. The vertical grey areas are clusters, which were classified as squat. See text for more details.

In Figure 2 an excerpt with 2 meter length of raw and processed inspection data of a single rail is shown.

3

UT B-scan data is in the top lane (#1), and red marks indicate regions with suspect UT feature data. UT features were derived from pixel depth values of individual channel types, and consecutively a feature reduction by Principal Component Analysis was performed.

ET complex plane raw voltage signals of 4 probes are given in the following lanes (#2-#5), and thereunder converted probe signals (lane #6). Red and magenta marked peaks were used as ET feature input.

In the lower half, the VT raw image (lane #7) and the cut and slightly processed image with results of Blob detection (#8) are shown. Blob parameters together with mean intensity peak changes (lane #9) were used as VT feature input.

The bottom lane (#10) summarizes various information: the squat reference positions (black lines), the clustering of the classification feature input, i.e. each 70 mm a block, their colour coding indicates the number of significant features for single system evaluation (blue – 0, green – 1, yellow – 2, red – 3). The vertical grey areas across all lanes indicate the clusters, which were classified as squat by usage of all systems' feature inputs.

For the labelling as classifier input, no distinction in sizing or some kind of severity classes was made, i.e. binary classification – 'squat yes or no?' – was done. About 20 meter section of one test run was used to train the model, whereas the corresponding section of the second test run was used to obtain the classifier test results.

In Table 1 classification measures are given. The TP, FP, FN and TN counts are binary contingency table entries, and the given rate and other performance measures are calculated from these values – see e.g. (10) for definitions. For each system individual classification results are given, and the results of the Combined Systems Method.
For CSM hit rate or sensitivity (TPR) was 97%, and specificity $(1-\mathrm{FPR})$ was 97%. Precision, i.e. TP over all positive predictions (TP+FP), for detection of given squat population was 78%. It can be seen that the combined evaluation provides a significantly improved hit rate with about the same or better specificity, if compared with single system results. This performance gain can be also seen with the likelihood ratios given in the table.

Table 1: Performance of binary classification of squats for single systems and Combined Systems Method (TP – true positives, FP – false positives, TN – true negatives, FN – false negatives, LR+ – positive likelihood ratio, LR− – negative likelihood ratio, TPR – TP rate, others accordingly)

performance measures		UT	ET	VT	Combined Systems Method
TP	counts	18	25	25	29
	rate	60%	83%	83%	97%
FP	counts	26	11	9	8
	rate	9%	4%	3%	3%
TN	counts	280	295	297	298
	rate	92%	96%	97%	97%
FN	counts	12	5	5	1
	rate	40%	17%	17%	3%
LR+	(TPR/FPR)	7.1	23.2	28.3	37.0
LR−	(FNR/TNR)	0.4	0.2	0.2	0.03

4. Conclusions and Outlook

Supervised machine learning technique for classification was applied on clustered feature data, either per individual inspection system (UT, ET or VT), or in combination, which we called 'Combined Systems Method'. The benefit of the CSM for squat classification in terms of true positive rate (TPR) and false positive rate (FPR) – as given in Table 1 – in comparison to the feature evaluation per single system is evident.

Because of the limited defect population of about 'mid-size' and severe squat defects for training and testing, the given performance measures are to be seen as indicative values. Binary classification results are given herein, but extension of the used classifier to multiple defect classes is in general straightforward with extended labelling. The underlying feature extraction for UT and ET data, respectively, is applicable for several other object or defect classes. For VT image evaluation for other defect types an extended or altered feature set, e.g. based on Convolutional Neural Network layers, would be needed, which could be used as improved input for the final CSM classification step.

Very recently a new ET system with 8 probes per rail was introduced (11), which will give optimized coverage for both, head checks at gauge corner, and squat-like or other RCF types, across whole wheel-rail contact zone, and therefore also an extended feature set as classification input.

Axle box acceleration (ABA) measurements were performed by Delft University of Technology (Netherlands) in the past with a system installed on UST96 inspection train of EURAILSCOUT. In (12) this ABA data was evaluated as promising candidate for detection of squat defects, even of small size. Data features of such recorded short scale surface geometry changes are seen as another input source for CSM.

References

1. M Papaelias, C Roberts and CL Davis , "A review on non-destructive evaluation of rails: State-of-the-art and future development", Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2008.

2. EE Magel, "Rolling contact fatigue: a comprehensive review", Federal Railroad Administration. Office of Railroad Policy and Development, United States, 2011.

3. Q Li, S Ren , "A real-time visual inspection system for discrete surface defects of rail heads", IEEE Transactions on Instrumentation and Measurement 61(8), 2189-2199, 2012.

4. P Bojarczak , "Visual algorithms for automatic detection of squat flaws in railway rails", Insight, 2013.

5. A Jamshidi, S Faghih-Roohi, S Hajizadeh, A Núñez, R Babuska, R Dollevoet, Z Li and B De Schutter , "A big data analysis approach for rail failure risk assessment", Risk Analysis 37(8), 1495-1507, 2017.

6. T Szugs, A Krüger, G Jansen, B Beltman, S Gao, H Mühmel and R Ahlbrink, "Combination of Ultrasonic and Eddy Current Testing with Imaging for Characterization of Rolling Contact Fatigue", in 19th World Conference on Non-Destructive Testing (WCNDT), Munich, 2016.

7. "PyData", available at: https://pydata.org/.

8. "OpenCV", available at: https://opencv.org/.

9. "scikit-learn: SGDClassifier", available at: http://scikit-learn.org/0.18/modules/generated/sklearn.linear_model.SGDClassifier.html.

10. DMW Powers, "Evaluation: from Precistion, Recall and F-measure to ROC, Informedness, markedness and correlation", Journal of Machine Learning Technologies, 2011.

11. R Ahlbrink, D Beilken, S Rühe and T Szugs, "Erweiterung der Wirbelstromprüfung zur Erfassung der gesamten Rollkontaktfläche", in ZfP im Eisenbahnwesen, Wittenberge, 2018.

12. M Molodova, "Detection of early squats by axle box accelerations", PhD thesis, TU Delft, 2012.