First published in:

Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008

HMM Based Segmentation of Continuous Eddy Current Sensor Signals

Stefan Hensel
Institut für Mess- und Regelungstechnik
Universität Karlsruhe
D-76131 Karlsruhe, Germany
hensel@mrt.uka.de

Carsten Hasberg
Institut für Mess- und Regelungstechnik
Universität Karlsruhe
D-76131 Karlsruhe, Germany
hasberg@mrt.uka.de

***Abstract*— An exact localization of trains is essential for effective disposition and design of modern train operating systems, allowing a better use of the given infrastructure. In this paper we propose to use turnouts on rail tracks as absolute landmarks and re-calibration points for onboard location systems. The measurements base on an eddy current sensor system, additionally providing speed information through correlating inhomogeneities along the rail track. This paper presents a hidden Markov model approach that offers a robust detection and separation of turnouts. The proposed algorithm makes it possible to process whole train stations with successive turnouts continuously, to perform a first low-level classification and to separate close events that can be accurately cut out of the signal, which is a basis for an advanced classification.**

I. Introduction

A robust and reliable positioning system is essential for modern and yet to come train operating concepts. Current approaches heavily rely on navigation beacons placed along the track, which are expensive in installation and maintenance. These landmarks fulfil the task to re-calibrate relative measurement systems, e.g. odometers, and provide an exact position even in dense vegetation and valleys where satellite systems tend to fail.

Recent approaches ([1],[2]) propose an onboard location system based on a combination of velocity measurements, digital map, and GPS information. The whole system is built upon a new eddy current sensor system which is capable of a non-contact speed measurement, not sensitive towards hard weather condition or wheel slippage. Besides calculating the speed information, which is achieved by correlation principles, the signals are applicable for turnout classification. The detection of the turnouts in the signal is at present implemented with several thresholds, whereas the classification is performed by a cross correlation with reference signals [3]. While the reported results indicate the general feasibility of this approach, the usability of the system suffered from the low automation level - the thresholds and parameters must be adapted manually to each track - and the fact that this method is incapable to separate events that follow in short distance within the signal. This disqualifies the system to cope with unknown tracks or explicitly separating and clipping turnouts from other occurrences for a later classification.

Hidden Markov models (HMMs) are capable to model a variety of nondeterministic systems and signals by assuming an unobservable Markov chain that governs the no stationary nature of the signal. Therefore, they are an ideal tool to solve the detection and classification of turnouts on the one hand and other track installations on the other hand in one step. We present in this paper a solution for the detection and segmentation problem of current systems and implement a stochastic approach, which can easily be augmented to a high-level classification system.

The remainder of this paper is organized as follows: A description of the employed eddy current sensor system is given first. Section III describes turnouts in general and introduces a signal model for the source and shape of the generated signals. HMMs and their implementation with the given system model are depicted in Section IV. Results and possible further applications are outlined in Section V, whereby the paper is concluded in Section VI with a prospect for further work.

II. Eddy Current Sensor System (ECS)

Eddy current sensors are used to detect inhomogeneities in the magnetic resistance of ferromagnetic materials. The given system consists of two differential sensors in a row, separated by the distance l. The sensors are placed in a housing for electromagnetical shielding, which is mounted 100 mm above the railhead of the train bogie. This enables the sensor to detect all major changes in the magnetic field along the track, mainly rail clamps but also turnouts and their components. Figure 1 shows the sensor system installed on a test train. A closed loop correlator [4] is used to determine the speed given the distance l. A typical signal $s(t)$ of the sensor system is shown in Figure 2. On open

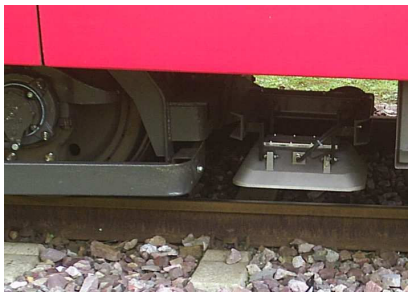

Fig. 1: Eddy current sensor system mounted on a railcar bogie.

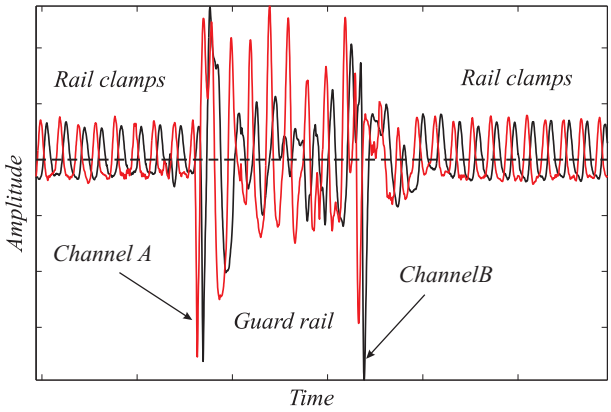

Fig. 2: Qualitative sensor signal $s(t)$ of the sensor signal for a guard rail amidst rail clamps.

tracks, the signal is mainly induced by the clamps whose equidistant spacing yields almost periodicity in the signal. A considerable change of the signal is observable when the sensor enters turnout areas. The amplitude is higher and changes notably which causes a change in the periodicity of the signal. Turnouts consist of many different parts, e.g. signal boxes, but the three main components present in every turnout are switchblade, frog and guard rail (see Subsection III-A).

The signals used for this paper are recorded with a sensor mounted on a tram of the local train operator. Test drives are carried out on a track situated near Karlsruhe, offering a length of approximately 18 km and 54 turnouts of which 21 are passed on a normal run.

III. Signal Model

For the detection and classification of turnouts with HMMs a specific signal model is needed. This section will first describe and explain turnouts as well as their different components. Proper features to separate the several components in the signal space are then derived and afterwards combined with an adequate signal processing and a source model for the signal.

A. Turnouts

Figure 3 shows a sketch of a turnout and the signal $s(x)$ produced by its components. The turnout can be subdivided into six segments that can be recognized in the signal. Due to different rail clamps, used for wooden sleepers inside the turnout area, segment I exhibits higher amplitudes. The subsequent switchblade segment II includes high peaks at the welding points before the movable part of the switchblade starts. Segment III is characterized by an area with higher amplitude than common turnout rail clamps (IV) can induce because of the higher amount of metal parts in that section, necessary to move the switchblade uniformly. The fifth segment contains either frog or guard rail, depending on whether the train follows the left main track or the right branch track. The sixth and last area is again characterized by the amplitude of turnout rail clamps. These six sections form a characteristic chain of events when traversed by a train.

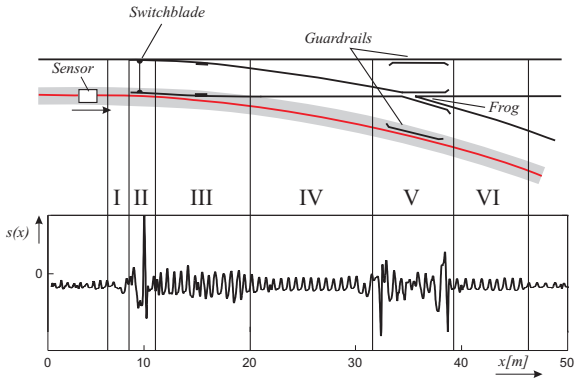

Fig. 3: Scheme of turnout and sensor signal $s(x)$ when driving on right branch track.

B. Classification Features

For each of the four directions a train can take when passing over turnouts, a signal sequence is observed that is characteristic for the individual components passed. It turned out that for a given sequence the major components are sufficient to know the direction: Detecting a guard rail after the switchblade is sufficient to know the train passes right, while the sequence switchblade and frog indicates a left turn.

Therefore, the signal model of the main components and possible disturbances, such as road crossings or rail joints, have to exhibit a distinct separability given the sensor signal. The component length and amplitude are convenient features according to the sensor system. The separation could not be improved with the frequency information of the clamps according to the fact that the rail clamps exhibit the same intrinsic distance in frogs and guard rails.

All guard rails on the test track are of the same type and have a length of 5 m. Frogs differ slightly in their length and shape depending on the turnout type and size (e.g. wye turnouts). They vary from 2 m up to 2.8 m. The switchblade model contains a strong amplitude peak at the beginning with a length of 1 m. The amplitudes depend on the amplification and have a ratio of approximately 1:0.4 for switchblade and guard rail and 1:0.5 given switchblade and frog. Possible disturbances that occur in the track are modeled with a length of 1 m to 1.5 m and lower amplitude than the main components. Figure 4 shows a three dimensional sight of the three main components with their particular normal distributions and the assumed disturbance model. The applied model assumes amplitude and length as normally distributed independent random variables A_i and L_i with $A_i, L_i \sim \mathcal{N}_{A,L}(\mu_i, \sigma_i^2)$, where i denotes the specific components. This assumption is appropriate due to bogie vibrations and shifts while passing the turnouts, different types of them and a slightly different track geometry at every turnout. All these influences are assumed to be statistically independent and gaussian distributed.

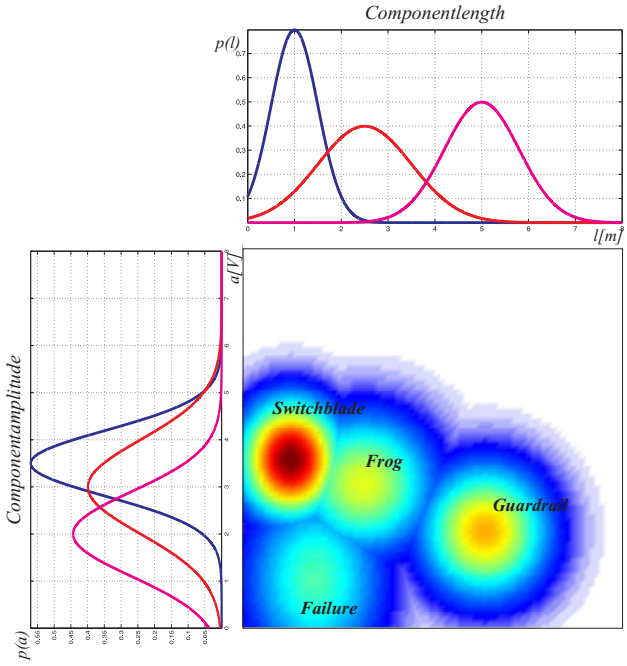

Fig. 4: Gaussian model for main turnout components and disturbance.

C. Signal Preprocessing

In order to use the statistical model described in Subsection III-B a rectification of the signals is conducted first.

The signal is filtered with a band pass filter to remove a possible offset and high frequency noise. To use the component lengths as a feature requires a transformation of the signal $s(t)$ into a speed independent spatial signal $s(x)$. This is achieved by a combination of the raw time signals and the respective speed output of the closed loop correlator. The transformation adds another source of stochastic jitter caused by uncertainties in the speed estimation, which originates from the cross correlation as well as from sensor movements of the bogie (see [5]).

A common approach for the analysis of oscillating signals in speech and signal processing is to examine the signal envelope [6]. The algorithm applied in this paper is composed of two steps. First the power signal is computed by short-term integration of the squared signal over the integration length T. A moving average (MA) filter for smoothing is applied afterwards. The resulting signal resembles to the low pass filtered magnitude of the Hilbert transform and represents the smoothed envelope of the original signal, which is the basis for subsequent pattern matching algorithms. Figure 5 shows the raw signal and the envelope later used for the pattern matching algorithms. It illustrates that the adoption of the preprocessing steps result in an higher turnout component length in spatial space and a reduced spatial resolution. Symmetric filters ascertain invariance of the modes to this processing.

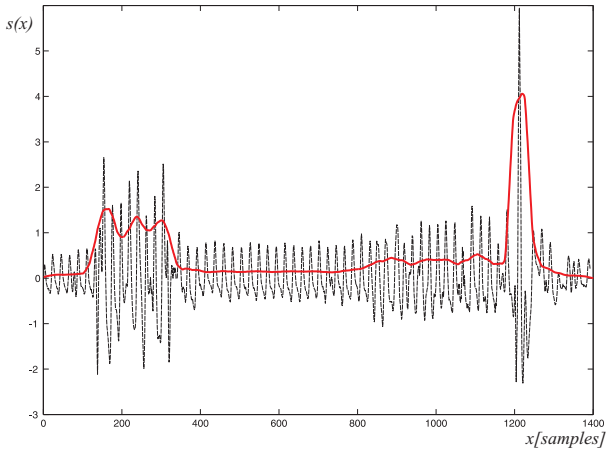

Fig. 5: Envelope of the signal after preprocessing.

D. Signal Generation

The discussion in Subsection III-B and III-C was focused on the analytical properties of the signal. The signal generation process is probabilistically modeled using the following two assumptions:

- Turnouts appear randomly distributed in the signal (presuming no map is present).
- Each component of the turnout emits a specific signal sample which is drawn from its underlying distribution.

The sensor signal is consequently interpreted as a two-stage stochastic process. In the first stage, a component class Ω_k is selected from the distribution p_k over all possible classes $k \in \{1..K\}$. The second stage is modeled by a multivariate continuous process with the conditional probability density $P(\mathbf{x}|\Omega_k)$. This property reflects the fact that even the same turnout can produce different signals depending on the speed measurement quality and other random influences.

With this model for the generation and the analytical shape of the signal it is possible to implement a sophisticated stochastic model.

IV. Hidden Markov Models

Hidden Markov models (HMMs) are state of the art statistical models used in a large variety of applications, such as time series in automated speech recognition and apply widely in biological sequence analysis [7], handwriting recognition [8] and simulation of stochastic processes [9]. A good introduction to HMMs can be found in [10] and [11], a comprehensive paper with an excellent tutorial is [12].

A. HMM Theory

A Markov process is a sequence of states $\mathbf{q} = q_1, q_2, ...,$ which take on values out of a set of elements $Q = \{s_1, ..., s_N\}$. In a first order Markov process, the conditional probability of state q_t at time instant t given all prior states is completely determined by the previous state q_{t-1}, i.e. $P(q_t|q_1...q_{t-1}) = P(q_t|q_{t-1})$. All possible state transitions are comprised in the so called *transition matrix* $\mathbf{A} = [a_{ij}]_{N\times N}$ according to $a_{ij} = P(q_t = s_j|q_{t-1} =$

s_i). Given the matrix $\mathbf{A}$ and an initial vector $\boldsymbol{\pi}$, obeying $\sum_j a_{ij} = 1$ and $\sum_{i=1}^{N} \pi_i = 1$, a so called Markov chain is completely described.

In addition to the Markov chain, a second stochastic process is generates a symbol of the set $V = \{v_1, ..., v_K\}$ at every time step. For an observer of the process only the emitted discrete series of symbols is visible, defined as $\mathbf{O} = O_1, O_2, ..., \ O_t \in V$ which depends on the states taken at every time step and is expressed as the probability $P(O_t|q_t)$. This leads straightforward to the definition of the so-called *emission matrix* $\mathbf{B} = [b_{jk}]_{N \times K}$ where $b_{jk} = b_j(v_k) = P(O_t = v_k|q_t = s_j)$ and $\sum_j b_{jk} = 1$.

We are thus left with a two-stage stochastic process where the hidden (non-observable) state sequence $\mathbf{q}$ induces the observation sequence $\mathbf{O}$. Every HMM is therefore completely determined by its N states, the possible symbols K and the parameter set $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \mathbf{A}, \mathbf{B})$. For the later implementation it should be mentioned that continuous probability density functions are used for the emissions according to $P(O_t = \mathbf{x}|q_t = j) = b_j(\mathbf{x})$ where $b_j(\cdot)$ obeys $\int_x b_j(\mathbf{x}) d\mathbf{x} = 1$. This accounts for the specific continuous features as used in the remainder of this paper. With a fully specified HMM the probability of the non-observable state given the emitted symbols can be estimated through computing its posterior:

$$P(\mathbf{q}|\mathbf{O}) \propto P(\mathbf{O}|\mathbf{q})P(\mathbf{q}) \propto \prod_{t=1}^{T} p(O_t|q_t)\, p(q_1) \prod_{t=2}^{T} p(q_t|q_{t-1})$$

B. HMM-based modeling of the ECS Signals

As explained before, each turnout is varying in dependency of its type, specific geometric attributes and non predictable noise. In Section III a stochastic signal model for turnouts has been introduced which fits perfectly to the HMM approach presented in this section. In Section III-B we pointed out, that the main information given in each sensor signal is amplitude and length of the singular turnout parts. The following subsections explain the implementation of each feature into the HMM framework. The adaption of the model parameters according to analogue ECS signals is thereby called training to the common machine learning vocabulary.

1) Length and transition matrix: Each turnout can be subdivided into six areas as shown in Figure 3. After the transformation into the spatial space the length of each segment is equal when passing the same turnout. Variations may only occur due to uncertainty in speed estimation. The first step is to assign each area of the turnout to a state of an HMM.

The arrangement of states and transitions depends on the application. Therefore many different HMM architectures are described in literature. A model architecture or model topology is derived by sketching the HMM as a finite state machine. If the underlying transition matrix $\mathbf{A}$ is fully occupied the model is often called ergodic, while for the modeling of time series left-to-right models are applied [12]. Figure 6 displays the left-to-right model fitted into the turnout scheme. An advantage of this model is the sparsity of the transition matrix and explicit start and end points of the Markov chains. It is now possible to model the length of

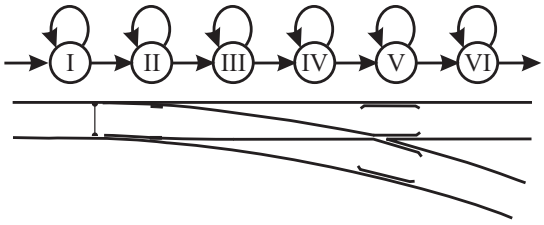

Fig. 6: Left-to-right model applied on a single turnout.

each area as a state occupancy duration of an HMM state. A limitation of the underlying Markov chain is the fact that the probability for the duration d to stay in a state i is geometrically distributed according to $P(d) = a_{ii}^{d-1}(1-a_{ii})$. Therefore the average state occupancy for a known self transition probability a_{ii} is given to $\overline{d} = \frac{1}{1-a_{ii}}$.

The key components to distinguish between a left and right turn of the train are the frog and the guard rail. The analysis of the sensor signals and the assumed model displayed in Figure 4 indicates that due to widely overlapping distributions of the amplitudes, a geometric distribution model is insufficient to separate the components.

Hence, the proposed method adopts state tying, where multiple states share the same observation distribution. Each single state is replaced by l new states with identical self transition probability. This results in a new state occupancy distribution, which is equivalent to that of a random variable consisting of the sum of l independent geometrically distributed random variables. The distribution of this sum is again a discrete distribution, the negative binomial distribution which has a none-zero mode. The resulting probability mass function is given as

$$P((X) = n) = \binom{n-1}{l-1} a_{ii}^{l}(1-a_{ii})^{n-l}$$

where a_{ii} is the self transition probability and n is the discretized duration. The choice of l can be freely done for each state under the constraint that the mode is getting sharper with bigger l while the computational complexity for the HMM conditions to $O(l^2)$. After an appropriate l and n are chosen for each turnout component, the self transitions for each tied state in $\mathbf{A}$ can be calculated. The parallel implementation of two chains results in a bimodal distribution, which is applied in the turnout area IV, which length is either 13 m or 19 m depending on the turnout radius.

2) Amplitude and emission matrix: While the component length is expressed in the transition matrix $\mathbf{A}$, the amplitude information is modeled through the emission matrix $\mathbf{B}$.

Five HMMs $\boldsymbol{\lambda}_{i=1..5}$, are necessary to model the four possible crossings of a turnout plus the possible disturbances. Due to preprocessing steps and component variations are better results achieved if the models are trained on real data instead of assuming ideal behavior of length and amplitude. Four different turnouts crossed in a different direction have

been selected, and the sections have been labeled manually. The emission probabilities of each state are assumed to be normal distributed according to $b_j(\mathbf{x}) \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Therefore, the parameters $\theta = (\mu_j, \sigma_j^2)$ can be determined with a maximum likelihood estimator based on the amplitude of each sample x_i of every state j according to $L(\theta_j|X) = \prod_{i=1}^{N} P(X = x_i|\theta_j)$ [13].

3) Continuous recognition: A major problem of recent approaches is the separability of the single events, e.g. two turnouts following in close dstance. The following approach solves this problem in two steps: After the models for each driving direction are trained, it is possible to separate turnouts from other events and classify the driving direction. Therefore a single turnout is classified by choosing the model which maximizes the probability $P(\mathbf{O}|\boldsymbol{\lambda}) = \sum_{j=1}^{N} \alpha_t(j)\beta_t(j)$, where α_t and β_t are the forward and backward probabilities for being in a state j at the time t (see [12], [7] for further details). In succession, the more interesting task is to give the most probable state sequence for several events in a given area, e.g. when passing a train station. Hence, secondly, an additional state for normal clamps is introduced which connects the individual events. The connection of the several models through a so called glue state enables detection and classification of arbitrary events. The topology for this continuous detection task is depicted in Figure 7. The task of finding the most probable

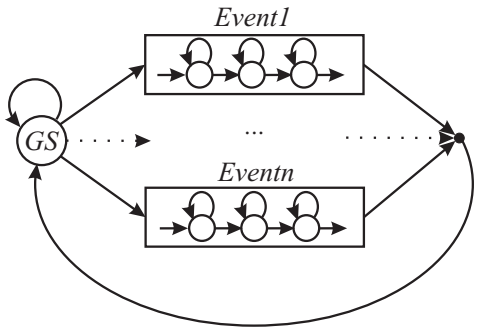

Fig. 7: Topology for a continuous event detection with glue state (GS).

state sequence for given observations is accomplished by the Viterbi algorithm [14].

V. Results

For the performance test of the proposed HMM a model was built up as explained in the previous sections. Four turnout directions and a disturbance model were expressed by 166 states. A first verification of the model was done by viewing the HMM not as an stochastic acceptor but as a generator. The shape and duration of the generated turnouts are exemplarily shown in Fig. 8 and fit well in length and amplitude when compared to real signals. The variance in the state duration was significantly improved, what was proved by simulating the generation of state transition with a desired duration. The variance of this transition, generated with the standard geometric distribution is 95% of the mean, compared to 45% of the mean when modeled with the

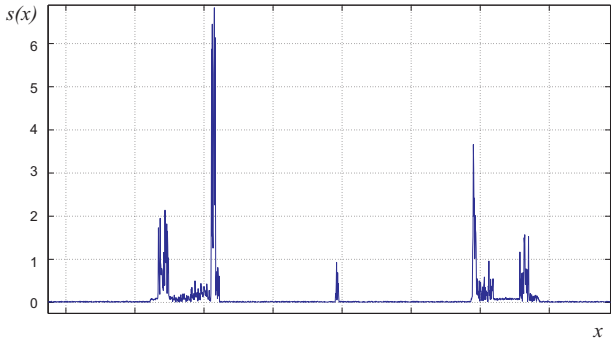

Fig. 8: Simulated turnouts with HMM as generator.

negative binomial model described in Section IV-B.1. For the evaluation of the most probable state sequence the Viterbi algorithm was implemented. The HMM parameters are adapted from real ECS data preprocessed as described in III-C down sampled by a factor eight. Therefore each sample in the following figures represents 16 cm of the track. The test sequences were taken with various length and containing various events to test several situations. The HMM output is marked as a solid red line. The height h of this line indicates the classification output of the distinct events: 0 encodes the common rail track areas with normal clamps, 1,2,3 and 4 are chosen for the four possible turnout passing directions, 5 indicates a disturbance. Figure 9 exemplarily shows results in separating the two turnouts and the disturbances, in the given case a road crossing. The considered sequence has a length of 136 m. Although they are extremely close, the turnouts are cut out very sharp and can be separated, which is not possible with conventional threshold methods. Besides employing the

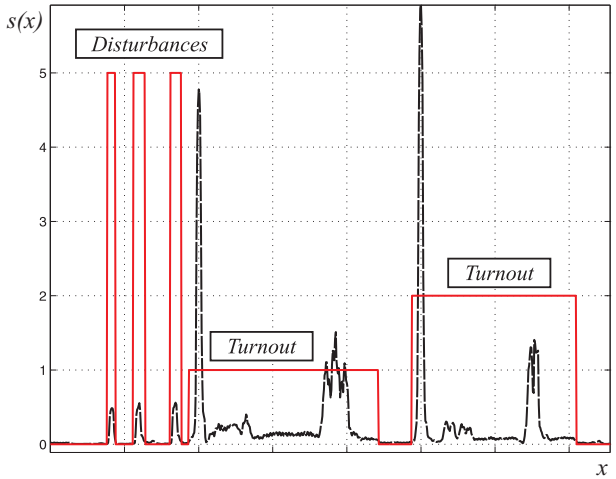

Fig. 9: HMM output and low passed envelope of the test sequence illustrating separation of single events.

HMM for the separation task which exhibits an excellent performance it is possible to use the output of the HMM as a classifier. The difficulty here is the intra class variance for components of the same type. Figure 10 shows the results for three turnouts in a train station, whereas turnout 2 and 3 are separated by clamps diverse of standard parts that are

correctly classified as a disturbance. While turnout 2 and 3

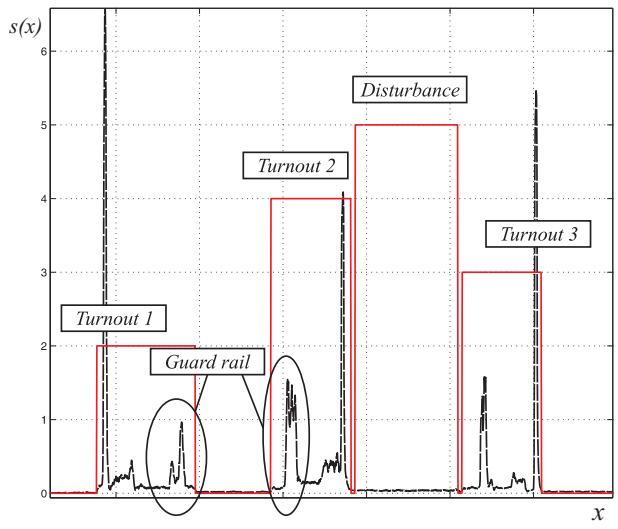

Fig. 10: Turnout detection with wrong classification.

are classified correctly, the misclassification of turnout 1 is a result of the deviation in the amplitude of the guard rail compared to the one used to train the model.

One of the main advantages of the given model architecture is the easy augmentation; new events like turnout crossings or new turnout types are easily implemented and trained. This paper outlines how to detect driving directions in a first low-level classification with as less states as possible to assure a good generalization. Therefore, more complex and specialized models should be applied in later stages to reduce computational complexity in the detection and cut-out stage.

A more simple way to rectify the result is to check the length of the components after the classification with the help of the state output. Frogs longer than 3.5 m and guard rails shorter than 3.5 m are detected and reclassified.

The most effective and thus proposed method is to use the given HMM as a detector which can give also an a priori probability of the direction. The detected segment is cut out and then classified by a more specialized HMM, e.g. for finding specific turnouts.

In our experiments all 20 turnouts were detected and cut-out correctly. True positive classification rate, i.e. rightly classify the driving direction of the train, was 61.5%. A false negative, i.e. a turnout classified as disturbance or not detected, never occurred at all.

VI. Conclusions and Future Work

A. Conclusions

This paper proposes a novel stochastic approach to detect turnouts in noisy signals. The proposed HMM approach fits well to the described signal generation and occurring difficulties. The HMMs easily outperform the hitherto used methods for detection and separation. They provide a high potential for automation and hierarchical classification. Disturbances along the track, stochastic variances of the signal caused by vibration or electro magnetic jitter as well as systematic failures that depend on uncertain speed measurements are successfully overcome. Available a priori knowledge, e.g. a distribution of the turnouts according to a poisson distribution, is directly implementable in the presented framework. The applied Viterbi algorithm allows a continuous detection with a recursive correction of the best path for every new sample. The calculation time for a sequence length of 500 m is currently 6.74 s when implemented in Matlab.

B. Future Works

The classification rate may be further improved through expanding the HMM by new models for special turnouts or characteristic components. All current misclassifications are based on the similarity of guard rails and frogs. Although a negative binomial distribution is far better than the simple geometric HMM state duration distribution, it is still not sufficient in worst case scenarios. On the other hand it is still necessary to allow a variation in the length caused by the big variety of different turnout types. Further research will focus on this problem with more specialized models, methods and additional training data.

References

[1] F. Böhringer, "Train location based on fusion of satellite and train-borne sensor data," in *Location Services and Navigation Technologies* (Y. Zhao, H. Klotz Jr., and L. Stockum, eds.), vol. 5084, pp. 76–85, SPIE, Bellingham WA, 2003.

[2] A. Geistler and F. Böhringer, "Detection and classification of turnouts using eddy current sensors," in *Computers in Railways IX* (J. Allan, R. Hill, C. Brebbia, G. Sciutto, and S. Sone, eds.), (Southampton), pp. 467–476, WIT Press, 2004.

[3] F. Mesch, F. Puente León, and T. Engelberg, "Train-based location by detecting rail switches," in *Computers in Railways VII* (J. Allan, R. Hill, C. Brebbia, G. Sciutto, and S. Sone, eds.), (Southampton), pp. 1251–1260, WIT Press, 2000. ISBN: 1-85312-826-0.

[4] T. Engelberg and F. Mesch, "Eddy current sensor system for non-contact speed and distance measurement of rail vehicles," in *Computers in Railways VII* (J. Allan, R. Hill, C. Brebbia, G. Sciutto, and S. Sone, eds.), (Southampton), pp. 1261–1270, WIT Press, 2000.

[5] T. Engelberg, "Speed measurement of rail vehicles using shielded eddy current sensors – field-test 3/00 in munich," Tech. Rep. AV 05/00, Institut für Mess- und Regelungstechnik, Universität Karlsruhe (TH), 2000.

[6] A. Oppenheim and R. Schafer, *Discrete Time Signal Processing*. Prentice Hall, 2nd edition, 1999.

[7] R. Durbin, *Biological sequence analysis*. 7, Cambridge University Press, 2002.

[8] A. Kundu and L. Bahl, "Recognition of handwritten script: a hidden markov model based approach," in *International Conference on Acoustics, Speech and Signal Processing*, pp. 928–931, 1988.

[9] Y. Guedon, "Hidden hybrid markov/semi-markov chains," in *Computational statistics & Data analysis*, 2004.

[10] R. Duda, P. Hart, and D. Stork, *Pattern Classification - 2nd edition*. Wiley Interscience, 2001.

[11] C. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, 2006.

[12] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, 1989.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, New Jersey, 1993.

[14] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, pp. 260–269, 1967.