

One-Class Classification for Anomaly Detection in Wire Ropes with Gaussian Processes in a Few Lines of Code

Erik Rodner Esther-Sabrina Wacker Michael Kemmler Joachim Denzler

Chair for Computer Vision, Friedrich Schiller University of Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany

{erik.rodner,esther.wacker,michael.kemmler,joachim.denzler}@uni-jena.de

Abstract

Anomaly Detection in Wire Ropes is an important problem. Detecting suspicious anomalies in the rope surface is challenging because of the variety of its visual appearance caused by reflections or mud on the rope surface. This hinders the discrimination between uncritical variations and small defects within the rope surface enormously. The fact that nearly no defective samples are available to train a supervised system relates this problem to the concept of one-class classification (OCC). In this work we show how to utilize one-class classification with Gaussian processes (GP) to detect anomalies in wire ropes. The method allows modeling the distribution of non-defective data in a non-parametric manner. Furthermore, it is really easy to implement (few lines of code), embedded in a Bayesian framework, and can be used with arbitrary kernel functions. Therefore, it is suitable for a wide range of defect localization applications. Our experiments, performed on two real ropes, demonstrate that the GP framework for OCC clearly outperforms former approaches for anomaly detection in wire ropes. The obtained results are comparable with or even outperform those obtained with the Support Vector Data Description, which is the state of the art reference in the field of one-class classification.

1 Introduction

Wire ropes are used in more fields of daily life than one might think. Elevators, bridges and ropeways are just a few examples. For this reason, wire ropes have to be inspected regularly to ensure their reliability. The manual visual inspection of wire ropes is a challenging task. The heavy ropes cannot be unmounted and cleaned in advance. Furthermore, the inspection speed is quite high to ensure an analysis of long ropes in acceptable time. Additionally, the periodic structure of wire ropes contributes to the fact that a manual visual inspection is an exhausting and error-prone task.

The following two reasons complicate an automatic, visual inspection of wire ropes: at first, there are not enough defective samples available to model the arbitrary defect characteristics. For this reason, it is not possible to train a classification framework in a supervised manner. Secondly, the image data exhibits a high variance and a discrimination between a noisy appearance and a real defect is a non-trivial problem even for a human expert. Figure 1 gives an example for a typical surface defect in wire ropes.

One-class classification (OCC) [12], also known as outlier detection [4] or novelty detection [6], comprises the learning of a binary classifier just given a set of samples from a single target (or positive) class. The

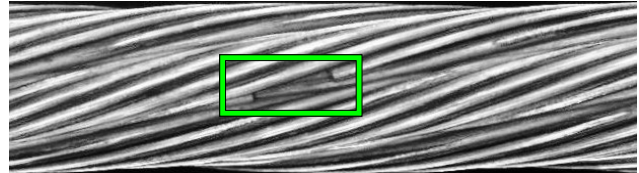


Figure 1. A typical surface defect: a broken wire

goal is to detect samples which are unlikely to belong to the target class. This concept is suitable for a problem like anomaly detection in wire ropes, where the number of defective samples is very limited.

There are various techniques to model the target class: one common approach is to model the distribution of the positive examples by parametric, generative models such as Gaussian mixture models [1]. Also boundary methods like the k-nearest neighbor classifier or reconstruction methods like k-means or principal component analysis can be involved to represent the target class [12]. Recently published work mostly uses kernel methods like the one-class Support Vector Machine (1-SVM) [11] or the highly related Support Vector Data Description (SVDD) [13], which use the kernel trick to model the data distribution in a non-parametric manner.

The application of Gaussian processes (GP) in machine learning also leads to a kernel based approach which can, in contrast to SVM related methods, be formulated in a Bayesian framework. A recent work of Kemmler *et al.* [5] shows how to use GP for OCC. In the following paper we demonstrate how to detect defects on wire ropes with this technique and emphasize its advantages, such as a simple implementation.

The remainder of this paper is structured as follows: in section 2 we review related work dealing with an automatic visual inspection of wire ropes. Section 3 introduces the usage of Gaussian process priors for OCC problems. In section 4 we explain how to perform anomaly detection in wire ropes. Our experiments compare the various OCC approaches in the context of anomaly and defect detection in wire ropes. They are presented in section 5. Finally, we conclude the paper with a discussion of our results.

2 Related Work on Wire Rope Analysis

Due to its relevance for rope safety, several approaches for an automatic visual inspection of wire ropes were developed in the past [8, 9, 3, 14]. All of them operate on image data, provided by a line camera system comparable to the one described by Moll [7].

All of the mentioned work has the common goal to detect anomalies in the measured rope data which ad-

vert to a possible defect. Whereas the approaches of Platzer *et al.* [8, 9] and Haase *et al.* [3] focus on the visual appearance of the rope, Wacker and Denzler [14] present a strategy for an image-based monitoring of important rope variables as the lay lengths of strands and wires. Their approach is purely focused on the regular rope structure and it allows a detection of creeping changes in these rope variables. However, they are not able to diagnose anomalies, which change the visual appearance of the rope surface as such as corrosion or broken wires.

The work of Platzer *et al.* and Haase *et al.* can be grouped into two different categories. In [9] the anomaly detection in wire ropes is performed by a generative Gaussian mixture model and the suitability of different textural features for this task is evaluated. The other category consists of approaches, which additionally incorporate the context imposed by the sequential character of the rope [8, 3].

Our intent is to show how well OCC can work in this scenario without using any context knowledge or structural information of wire ropes. This allows us to develop a generic method which is suitable for a wide range of other defect localization applications. We compare our results to those obtained by the generative Gaussian mixture model described by Platzer *et al.* [9], which also does not exploit context information.

3 One-Class Classification with GP

This section gives a brief introduction to one-class classification with Gaussian process priors. First we explain the basic ideas of machine learning with GP priors [10] which is followed by a description of their usage for one-class classification [5].

3.1 Gaussian Process Priors

Given n training examples $\mathbf{x}_i \in \mathbf{X} \subset \mathbb{R}^D$, which denote feature vectors, and corresponding binary labels $y_i \in \{-1, 1\}$, we would like to predict the label y_* of an unseen example \mathbf{x}_* . The goal in classification is to find the intrinsic relationship between inputs \mathbf{x} and labels y . It is often assumed that the desired mapping can be modeled by $y = f(\mathbf{x}) + \epsilon$, where f is an unknown function and ϵ denotes a noise term. One common modeling approach is to assume that f belongs to some parametric family $f(\mathbf{x}; \boldsymbol{\theta})$ and to learn the parameters $\boldsymbol{\theta}$ which best describe the training data.

However, the main benefit of the GP framework is the ability to model the underlying function f directly as a latent random variable, *i.e.* without any fixed parameterization (since all parameter configurations are taken into account). The posterior of the label y_* of an unseen example \mathbf{x}_* can be derived by marginalization of the latent function value $f_* = f(\mathbf{x}_*)$:

$$p(y_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int_{\mathbb{R}} p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) p(y_* | f_*) df_* \quad (1)$$

where we assume that the label is conditionally independent of the example if the corresponding function value is given. The function values \mathbf{f} of the training set \mathbf{X} are also latent leading to a second marginalization:

$$p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int_{\mathbb{R}^n} p(f_* | \mathbf{X}, \mathbf{f}, \mathbf{x}_*) p(\mathbf{f} | \mathbf{X}, \mathbf{y}) d\mathbf{f} \quad (2)$$

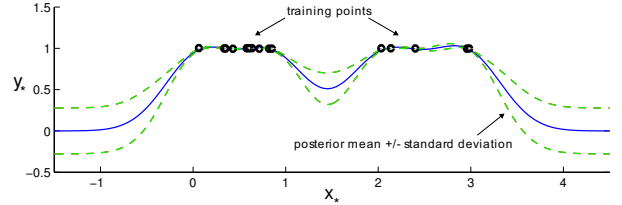


Figure 2. One-dimensional example of one-class classification with GP regression as proposed by [5]. The posterior mean and the negative standard deviation are both suitable OCC scores.

The prior of the latent function f can now be modeled as a Gaussian process $\mathcal{GP}(0, \mathcal{K})$ with zero mean and covariance function \mathcal{K} . This allows modeling the correlation of function values using the similarity of input examples calculated by a kernel function, like the radial basis function (rbf) kernel: $\mathcal{K}(\mathbf{x}, \mathbf{x}') = \exp(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}'\|^2)$. If we assume Gaussian noise ϵ , $p(y|f)$ is a Gaussian distribution and all involved marginalizations can be calculated in closed form leading to a posterior $p(y_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*)$ which is also Gaussian and has the following mean and standard deviation:

$$\mu_* = \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \quad (3)$$

$$\sigma_*^2 = k_{**} - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_* + \sigma_n^2 \quad (4)$$

\mathbf{K} represents the kernel matrix of the training data, \mathbf{k}_* denotes the vector of kernel values of the new example and the training set, $k_{**} = \mathcal{K}(\mathbf{x}_*, \mathbf{x}_*)$ and $\mathbf{y} \in \{-1, 1\}^n$ is the vector of all binary training labels.

3.2 Utilizing GP for One-Class Classification

In contrast to other supervised classification methods, the GP framework allows tackling the OCC problem directly. If we have only given n training examples \mathbf{x}_i , which all belong to a single (positive) class $y_i = 1$, the Bayesian formalism described in the previous section still holds and the inference leads to suitable solutions for OCC applications. This is due to the zero-mean GP prior on f which favors functions around zero. Without this prior the simplest explanation would be the function $y_* \equiv 1$ which is completely unsuitable for one-class classification. Figure 2 shows an example of GP regression applied to an one-dimensional example. It can be seen that the utilization of a GP prior leads to a posterior mean function which has high values (around $y = 1$) in high density areas next to the training points and decreases monotonically if the distance to the training set increases. Another important fact which is illustrated in Figure 2 is that the posterior variance shows an opposite behavior with high values in outlier regions. The consequence is that the posterior mean and the negative variance are suitable OCC measures, which is also validated in a more theoretical manner in [5].

The relation to other density estimation techniques becomes obvious if we consider the case of a single training point \mathbf{x} ($n = 1$) and the use of a rbf-kernel, which simplifies the formula for the posterior mean to:

$$\tilde{\mu}_* = \left(\frac{1}{1 + \sigma_n^2} \right) \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_* - \mathbf{x}\|^2 \right) \quad (5)$$

This is equivalent to an unnormalized normal distribution with mean value \mathbf{x} if we assume noise-free observations ($\sigma_n^2 = 0$). Other relations to *e.g.* Parzen density estimation or normal distributions in feature space can be found in [5]. The same work also studies other measures for OCC derived from the GP framework and approximation techniques when considering non-Gaussian noise. For our application to anomaly detection in wire ropes we only used the posterior mean and variance of GP regression, because all measures and variants showed a comparable performance in the image categorization experiments of [5].

3.3 Implementation

The implementation of GP for one-class classification is simple and straightforward in the regression case, especially for the posterior mean. First of all, the kernel matrix \mathbf{K} has to be computed with an arbitrary kernel function such as the rbf-kernel. The only thing which has to be done for training is to solve the linear equation system $(\mathbf{K} + \sigma_n^2 \mathbf{I}) \boldsymbol{\alpha} = \mathbf{y}$ with \mathbf{y} being an n -dimensional vector of ones. A solution can be found with the Cholesky decomposition, which involves $\mathcal{O}(n^3)$ operations. Afterwards, the estimated posterior mean is $\mathbf{k}_*^T \boldsymbol{\alpha}$ which involves $\mathcal{O}(n)$ operations. Calculating the posterior variance is similar but involves $\mathcal{O}(n^2)$ operations during testing.

4 Anomaly Detection in Wire Ropes

Our evaluation of one-class classification with GP priors is based on the anomaly detection problem which arises in the context of automatic visual rope inspection. In the following section we describe how to compute image features used for our approach as well as for the comparison with other OCC methods.

Due to the findings of Platzter *et al.* [9] we make use of histograms of oriented gradients (HOG) [2] for the description of the observed rope surface. HOG features are well-suited for the problem of anomaly detection in wire ropes and outperform most of the established statistical features, which are usually used for visual inspection tasks [15]. The regular structure of wire ropes exhibits articulated gradient orientations along the twist direction. Gradients with a perpendicular orientation usually can be considered as anomalous. The HOG descriptors are computed from gradient images. For this purpose, a sequence of 1d rope measurements is concatenated to a time frame comprising 20 camera lines. The resulting 2d image has a width of 20 and a height dependent on the rope diameter in pixels. Subsequently, gradient images for these frames are computed and divided into small cells of 20×20 pixel. For each of these cells a gradient orientation histogram is computed whose entries are weighted with the gradient magnitude. Finally, a feature vector which gives a description for the whole time frame is formed by concatenating the normalized cell histograms. The normalization is performed with respect to the whole time frame. We used histograms with four discrete orientation bins, as the distinct number of gradient orientations in intact rope data is limited. As proposed by Platzter *et al.* [9], we further compute the entropy of the discrete distribution resulting from each cell histogram, to improve a discrimination between system-

atic deviations in the orientation histograms and noise caused by reflections, mud or abrasion. The resulting feature vector has a dimension of $D = 5 \frac{h}{20}$ where h is the rope diameter in pixels. Note, that in a preprocessing step the rope is automatically segmented and h is chosen to be the maximum pixel diameter observed in a time range covering a full lay length of the rope.

For all different OCC approaches, a confidence value is computed, which defines the likelihood of the sample belonging to the target class. Nevertheless, an adjustment of the confidence belt around the target class cannot be performed without a set of outliers or anomalies belonging to the counter class. On the other hand, this is in general the case for OCC and it allows an application dependent fine tuning of the decision threshold.

5 Experiments

Experimental Data: In our experiments we use two different rope data sets. Both were acquired under realistic conditions by the prototype system described in [7]. In the following we refer to the two different data sets by the terms ROPE1 and ROPE2. ROPE1 has a length of approximately 1.3km, the shorter ROPE2 is 400m long. The resolution of the line cameras is known to be 0.1 mm/camera line.

For each data set a ground truth defect labeling was provided by a human expert. The distinct difference between the two rope data sets is the complexity of the comprised errors. ROPE1 contains more obvious defects whereas those contained in ROPE2 are often inconspicuous and small. We trained our OCC classifiers on a training set computed from a rope sequence of 100,000 camera lines (10m rope, 5000 training examples). This defect free rope region was proposed by a human expert. The evaluation was performed on the remaining rope sequence which contains all labeled defects.

Experimental Setting: The variance σ^2 of the rbf kernel was set to $e^{-2.5}$ in all experiments. Additional experiments showed that for our application the influence of this parameter setting on the results is negligible. The noise parameter σ_n^2 was automatically determined. We iteratively increased the value of σ_n^2 ($0, 10^{-8}, 10^{-7}, 10^{-6}, \dots$) until the Cholesky decomposition of the kernel matrix can be calculated ensuring its positive-definiteness. An important parameter of the SVDD method is the outlier fraction ν , which was also experimentally analyzed but without a significant difference in the results. Therefore ν was set to 0.1.

The results obtained for the different methods are displayed in Figure 3 using ROC curves with the area under the ROC curve (AUC) given in the legend. Note, that the ROC curves are averaged over the results obtained for the four individual camera views.

Evaluation: It is obvious that all three kernel-based OCC approaches, posterior mean and variance of GP regression and SVDD, outperform the classical GMM strategy proposed in [9]. From the AUC values it becomes clear, that the GP based OCC approaches offers a slightly better performance than SVDD. The posterior mean approach achieves the best results and is also faster than the posterior variance approach during testing (*cf.* section 3.3). Figure 4 shows some example detections of our algorithm, where wire parts with a posterior mean below a manually selected threshold are recognized as defects. Please

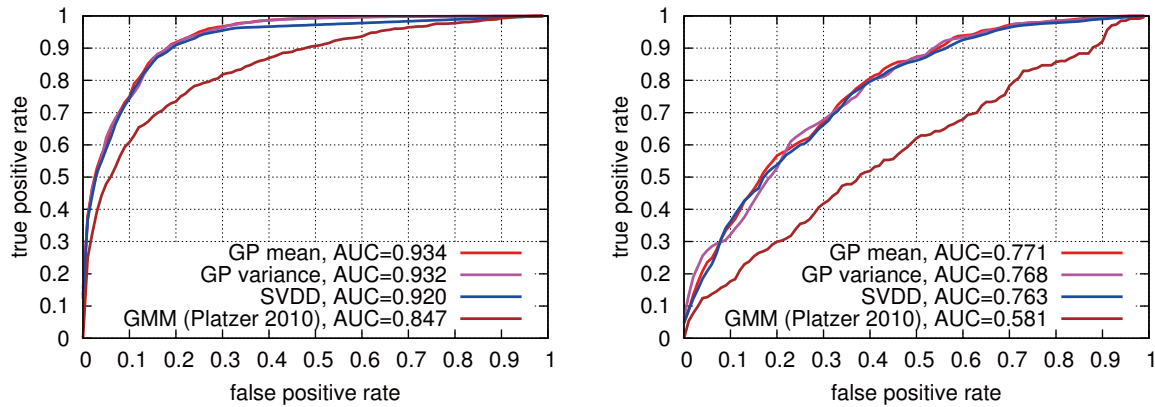


Figure 3. Average ROC curves for the (left) ROPE1 and the (right) ROPE2 dataset. The curves for all three kernel-based methods (GP mean, -variance and SVDD) are very similar and best viewed in color.

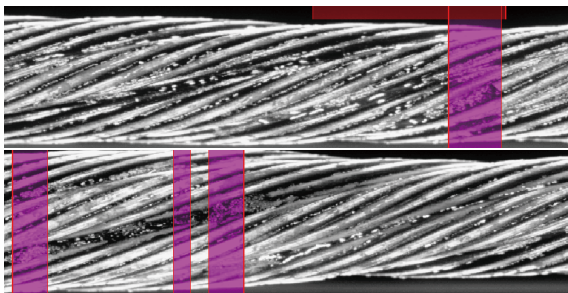


Figure 4. Example images of a correct defect detection (broken wire) and some false positives detected by our approach. Results are highlighted in magenta and the red bar on top of the rope shows the defect annotation of a human expert.

note that approaches which exploit the special structure of wire ropes achieve higher recognition results (e.g. [8] achieved an AUC value of 0.96), but we especially concentrate on defect localization without any prior knowledge to ensure an universal applicability with respect to other application areas.

6 Conclusions

We showed how to utilize one-class classification with Gaussian processes for defect localization in wire ropes. This kernel based approach outperforms previous OCC methods for defect localization significantly. Beyond that, it is very simple to implement and it is embedded in a probabilistic setting. Future research concentrates on the application to other defect localization problems and the development of specialized kernel functions which incorporate prior knowledge about the special structure of wire ropes.

References

- [1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.
- [3] D. Haase, E.-S. Wacker, E.-G. Schukat-Talamazzini, and J. Denzler. Analysis of Structural Dependencies for the Automatic Visual Inspection of Wire Ropes. In *Vision, Modeling and Visualization*, pages 49–56, 2010.
- [4] V. Hodge and J. Austin. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.
- [5] Michael Kemmler, Erik Rodner, and Joachim Denzler. One-class classification with gaussian processes. In *ACCV*, volume 2, pages 489–500, 2010.
- [6] M. Markou and S. Singh. Novelty detection: a review - part 1: statistical approaches. *Signal Processing*, 83(12):2481 – 2497, 2003.
- [7] D. Moll. Innovative procedure for visual rope inspection. *Lift Report*, 29(3):10–14, 2003.
- [8] E.-S. Platzer, J. Nägele, K.-H. Wehking, and J. Denzler. HMM-Based Defect Localization in Wire Ropes - A New Approach to Unusual Subsequence Recognition. In *DAGM*, pages 442–451, 2009.
- [9] E.-S. Platzer, H. Süße, J. Nägele, K.-H. Wehking, and J. Denzler. On the Suitability of Different Features for Anomaly Detection in Wire Ropes. In *Computer Vision, Imaging and Computer Graphics: Theory and Applications*, pages 296–308, 2010.
- [10] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- [11] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13:1443–1471, 2001.
- [12] D. M. J. Tax. *One-class classification - Concept-learning in the absence of counter-examples*. Phd thesis, Technische Universität Delft, 2001.
- [13] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Mach. Learn.*, 54:45–66, 2004.
- [14] E.-S. Wacker and J. Denzler. An Analysis-by-Synthesis Approach to Rope Condition Monitoring. In *International Symposium on Visual Computing*, pages 459–468, 2010.
- [15] X. Xie. A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques. *Electronic Letters on Computer Vision and Image Analysis*, 7(3):1–22, 2008.