



HAL
open science

Neighbor Retrieval Visualizer for Monitoring Lifting Cranes

Paul Honeine, Samira Mouzoun, Mario Eltabach

► **To cite this version:**

Paul Honeine, Samira Mouzoun, Mario Eltabach. Neighbor Retrieval Visualizer for Monitoring Lifting Cranes. Advances in Condition Monitoring of Machinery in Non-Stationary Operations: Proc. 6th International Conference on Condition Monitoring of Machinery in Non-stationary Operations, 2018, Santander, Spain. 10.1007/978-3-030-11220-2_3. hal-01965902

HAL Id: hal-01965902

<https://hal.science/hal-01965902v1>

Submitted on 27 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Neighbor Retrieval Visualizer for Monitoring Lifting Cranes

Paul Honeine¹, Samira Mouzoun¹, and Mario Eltabach²

¹ Normandie Université, Université de Rouen, LITIS (FR CNRS 3638),
Avenue de l'Université, 76800 Saint Etienne du Rouvray, France

`paul.honeine@univ-rouen.fr`

² Centre Technique des Industries Mécaniques, CETIM
52 av Félix Louat, 60304 Senlis, France

Abstract. Gear wear is hard to monitor in lifting cranes due to the difficulties to provide appropriate models of such complex systems with varying functioning modes. Statistical machine learning offers an elegant framework to circumvent these difficulties. This work explores recent advances in statistical machine learning to provide a data-driven model-free approach to monitor lifting cranes, by investigating a large number of indicators extracted from vibration signals. The principal contributions of this paper are twofold. Firstly, it explores the recently introduced Neighbor Retrieval Visualizer (NeRV) method for nonlinear information retrieval. The extracted information allows to construct a low-dimensional representation space that faithfully depicts the evolution of the system. Secondly, it proposes a simple and efficient detection method to detect abnormal evolution and abrupt changes of the system at hand, using the distance measure with neighborhood retrieval in the same spirit as NeRV. The relevance of the proposed methods, for visualizing the evolution and detecting abnormality, is demonstrated with experiments conducted on real data acquired on a lifting crane benchmark operating for almost two years with more than fifty indicators extracted from vibration signals.

Keywords: nonlinear information retrieval, neighbor retrieval visualizer, dimensionality reduction, gear wear monitoring, monitoring lifting cranes, detect abnormal evolution, abrupt change detection

1 Introduction

Early failure detection in machinery and industrial processes is essential to avoid abrupt disastrous breakdown. Many industrial facilities with motor-driven equipments are critical, and failures can harm equipments, environment, and personnel. This is the case in power plants, petrochemical industries, aerospace and military equipments [4, 13]. To assess the running conditions of motor-driven systems, non-invasive monitoring and diagnosis techniques have been increasingly investigated by researchers in both academic and industrial domains [16]. Vibration, acoustic emission, and supply current signals have been heavily explored to monitor such machines [17].

Crane winches are extensively used in most construction sites and many industrial applications, owing to the highly reliable electromechanical power [14, 2, 1]. Since early fault detection is essential to the safety of equipment and personnel, the assessment of the running conditions of a crane winch is crucial. Due to the reduced vibrational energy of gear wear, its detection is difficult as opposed to tooth crack or other localized faults. This paper presents a case study of a lifting crane winch, including an induction motor and a multistage epicyclic gearbox, with a dynamic load where an accelerated life test is conducted [5]. To this end, more than fifty “indicators” are extracted from the vibration signal, computed in both “up” and “down” operational modes of each cycle.

This paper explores recent advances in statistical machine learning to provide a data-driven model-free approach to monitor gear wear, by investigating multiple indicators extracted from several sensors installed on the lifting crane under study. For this purpose, the information retrieval is first studied in order to extract information that allows to represent, in a two-dimensional space, the data living in the high-dimensional space of more than fifty “indicators”. While conventional dimensionality reduction techniques, such as Principal Component Analysis (PCA) [12] and Multidimensional Scaling (MDS) [3], are not appropriate for nonlinear information retrieval as depicted in many research papers, this work explores the recently introduced Neighbor Retrieval Visualizer (NeRV) method for nonlinear information retrieval [18]. The second contribution of this paper is a novel unsupervised classification method to detect abnormal evolution and abrupt changes of the system at hand. To this end, a simple and efficient technique is proposed using the distance measure with neighborhood retrieval in the same spirit as NeRV. Finally, the relevance of the proposed methods, for information retrieval and abnormality detection, is established with experiments conducted on the aforementioned real data from the lifting crane benchmark. Comparative analysis is conducted with the state-of-the-art novelty detection with kernel principal component analysis (KPCA) [7].

The rest of the paper is organized as follows. The following section presents the investigated information retrieval method. The proposed novelty detection method is described in Section 3. Section 4 demonstrates the relevance of these methods with experimental results on real data extracted from the lifting crane winch under study. Section 5 concludes this paper.

2 Nonlinear Information Retrieval for Visualization

Consider N samples, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, living in an input space $\mathcal{X} \subset \mathbb{R}^d$. Information retrieval consists of extracting from each sample $\mathbf{x}_i \in \mathcal{X}$, for $i = 1, 2, \dots, N$, some relevant information represented by \mathbf{y}_i in some lower-dimensional space $\mathcal{Y} \subset \mathbb{R}^{d'}$, also called output space, with $d' \ll d$. When $d' = 2$, the output space can be viewed as a visualization of the data in a two-dimensional space, namely representing $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ with $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N \in \mathcal{Y}$. Therefore, information retrieval is related to dimensionality reduction.

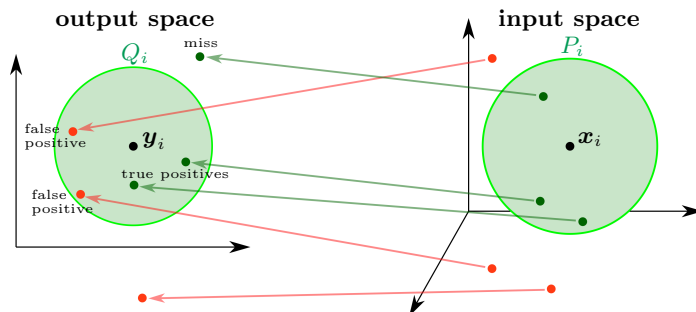


Fig. 1. Illustration of error types in information retrieval for visualization

Many dimensionality reduction methods have been proposed in the literature, since several criteria have been put forward by preserving some measure in both spaces. Conventional methods, such as MDS and PCA, rely on preserving the similarity or dissimilarity between pairs of samples in the input and output spaces, namely between $(\mathbf{x}_i, \mathbf{x}_j)$ and $(\mathbf{y}_i, \mathbf{y}_j)$ for all $i, j = 1, 2, \dots, N$. Many research papers have demonstrated that these methods are not suitable for nonlinear information retrieval. See for instance [6, 18] and references therein. As advocated in these references, the preservation of the neighborhood relationship as a similarity measure provide an appropriate nonlinear information retrieval. To this end, several methods have been recently introduced, the most known are Stochastic Neighbor Embedding (SNE) [6], with all its variants such as t-SNE, and Neighbor Retrieval Visualizer (NeRV) [18]. The latter is described in the remainder of this section.

A conventional approach to define neighborhood is the binary definition namely, for any $\mathbf{x}_i \in \mathcal{X}$ and its counterpart $\mathbf{y}_i \in \mathcal{Y}$, it determines which samples are in its neighborhood in the input and output spaces, respectively. Let P_i denote its neighborhood in the input space \mathcal{X} , namely the set of samples that are the closest to \mathbf{x}_i , and let $|P_i|$ be its cardinality, namely the number of neighboring samples. Analogously, let Q_i be the neighborhood of the sample \mathbf{y}_i in the output space \mathcal{Y} , and $|Q_i|$ its cardinality.

In the following, the optimization problem is defined, the goal being that $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ are determined such that the neighborhood of each \mathbf{y}_i is similar to the one of its counterpart \mathbf{x}_i in the input space \mathcal{X} . Let $N_{\text{TP},i}$ be the number of samples that are in both P_i and Q_i (*i.e.*, “true positives”), $N_{\text{FP},i}$ be the number of samples that are in Q_i but not in P_i (*i.e.*, “false positives”), and $N_{\text{miss},i}$ be the number of samples that are in P_i but not in Q_i (*i.e.*, “misses” or “false negatives”). These numbers are interrelated, since it is easy to see from Fig. 1 that $N_{\text{TP},i} = |P_i| - N_{\text{miss},i}$ and $N_{\text{TP},i} = |Q_i| - N_{\text{FP},i}$. In order to have a relevant information retrieval, one needs to minimize jointly $N_{\text{FP},i}$ and $N_{\text{miss},i}$ for all the samples. By considering the formalism from information retrieval and pattern recognition, the concepts of precision and recall are used in the

following. The precision (also called positive predictive value) is the proportion of relevant samples to the number of the retrieved samples, namely $N_{\text{TP},i}/|Q_i|$. The recall (also called sensitivity) is the proportion of relevant retrieved samples to the number of relevant samples, namely $N_{\text{TP},i}/|P_i|$. Finally, by averaging these two measures over all the samples with the expectation operator $\mathbb{E}(\cdot)$ over $i = 1, 2, \dots, N$, one gets the mean precision and mean recall to be maximized jointly.

The combination of the “binary” definition of the neighborhood with a hard thresholding and the aforementioned objective functions has several drawback. Indeed, the former does not take into account levels of neighborhood, and the latter do not penalize the violation of the ordering of neighbors. Moreover, the resolution of the two objective functions (mean precision and mean recall) is difficult to carry out since such optimization problems are ill-posed with an infinite number of solutions.

To overcome all these difficulties, a continuous definition of the neighborhood is considered in the following, by defining probabilistic neighborhoods in both input and output spaces. Given some $\mathbf{x}_i \in \mathcal{X}$, its neighborhood in terms of any \mathbf{x}_j is defined with the following probabilistic model of neighborhood:

$$p_{j|i} = \frac{\exp\left(-\frac{1}{\sigma_i}\|\mathbf{x}_i - \mathbf{x}_j\|_{\mathcal{X}}^2\right)}{\sum_{k \neq i} \exp\left(-\frac{1}{\sigma_i}\|\mathbf{x}_i - \mathbf{x}_k\|_{\mathcal{X}}^2\right)}. \quad (1)$$

Likewise for its counterpart $\mathbf{y}_i \in \mathcal{Y}$, the probabilistic model of neighborhood in the output space is defined as

$$q_{j|i} = \frac{\exp\left(-\frac{1}{\sigma_i}\|\mathbf{y}_i - \mathbf{y}_j\|_{\mathcal{Y}}^2\right)}{\sum_{k \neq i} \exp\left(-\frac{1}{\sigma_i}\|\mathbf{y}_i - \mathbf{y}_k\|_{\mathcal{Y}}^2\right)}. \quad (2)$$

In these expressions, $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$ denote the distances defined in the input and output spaces, respectively (without loss of generality, the Euclidean distance is considered in both spaces). The bandwidth parameter σ_i allows to control the scale of neighborhood of the i -th sample, with the same value set for both input and output spaces.

Since the neighborhoods in both input and output spaces are defined as distributions with a probabilistic model, it is natural to consider the Kullback-Leibler (KL) divergence, defined for the i -th sample by

$$D(p_i, q_i) = \sum_{\substack{j=1 \\ j \neq i}}^N p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad \text{and} \quad D(q_i, p_i) = \sum_{\substack{j=1 \\ j \neq i}}^N q_{j|i} \log \frac{q_{j|i}}{p_{j|i}}. \quad (3)$$

The KL divergences $D(p_i, q_i)$ and $D(q_i, p_i)$ are called respectively smoothed precision and smoothed recall (see Appendix A in [18] for connections to the aforementioned precision and recall definitions). Finally, averaging over all the samples, one gets the mean smoothed precision $\mathbb{E}(D(q_i, p_i))$ and the mean smoothed

recall $\mathbb{E}(D(p_i, q_i))$. It turns out that these two terms are antagonist in general. One elegant way to overcome this issue is the aggregated cost function using a convex combination, namely $\lambda\mathbb{E}(D(p_i, q_i)) + (1 - \lambda)\mathbb{E}(D(p_i, q_i))$, which is estimated over the set of N samples, as

$$\lambda \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} + (1 - \lambda) \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N q_{j|i} \log \frac{q_{j|i}}{p_{j|i}}.$$

The parameter $\lambda \in [0; 1]$ controls the tradeoff between mean smoothed precision and mean smoothed recall. When $\lambda = 1$, the SNE cost function is obtained [6]. The above cost function is minimized using a standard conjugate gradient algorithm.

3 Proposed Detection Method

This section introduces a novelty detection algorithm based on the distances and neighborhoods, thus in the same spirit as the method described in Section 2. For this purpose, the detection rule is applied on the extracted information. Since the “up” and “down” operational modes are naturally different, a detector is considered for each mode. For the sake of clarity and conciseness, one of the two detectors is described next.

To learn the detector and its parameters, the first N_{tr} samples are considered as training data, where N_{tr} is fixed by the experts under the condition that the samples were acquired from the system in “faultless” normal operating state. In our case, experts have determined that the system was in its normal operating state prior to cycle #25000. With some abuse of notation, let $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_{\text{tr}}}$ be the training samples for the mode under scrutiny (*i.e.*, “up” or “down”).

The main idea of the proposed detection method is to compute the distance of a given sample to the training samples, and compare it to a threshold in order to determine if the sample is abnormal or not. The detection technique consists of two stages. The first one is the training stage, where the parameters are estimated, namely the detection threshold. Based on the set of training samples, the threshold is set as follows

$$s = \max_{i=1, \dots, N_{\text{tr}}} \min_{\substack{j=1, \dots, N_{\text{tr}} \\ j \neq i}} \|\mathbf{y}_i - \mathbf{y}_j\|_{\mathcal{Y}}. \quad (4)$$

The second stage is the decision, namely detecting if a given sample is from an abnormal event or the normal operating state. When dealing with some sample \mathbf{y}_i , the decision rule consists in computing its distance to its nearest neighbor from the training samples, and comparing it to the above threshold, namely

$$\min_{j=1, \dots, N_{\text{tr}}} \|\mathbf{y}_i - \mathbf{y}_j\|_{\mathcal{Y}} \stackrel{\text{abnormal}}{\underset{\text{normal}}{\geq}} s. \quad (5)$$

In order to reduce the false alarm, the temporal information is considered by setting a sliding window as follows: an abnormal event is detected if three consecutive samples have distances that exceed the fixed threshold.

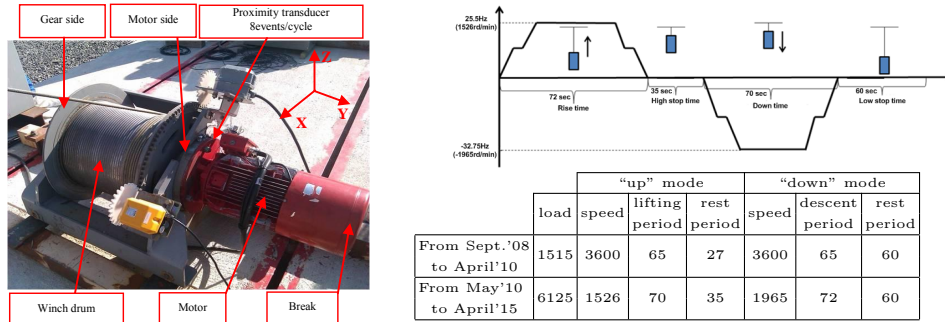


Fig. 2. Left: Overview of the lifting crane winch. Top-right: Winch operation scheme, with an “up” mode and a “down” mode interlaced with rest periods. Bottom-right: The two settings considered in the accelerated life test, with the load (in kg), the speed (in rpm) and the living and rest periods (in seconds) for both modes.

4 Experiments on the Lifting Crane Winch

This section provides experimental results to assess the relevance of the proposed methods. But before, it presents the benchmark of the lifting crane winch, its operational modes, and the extracted indicators/signals.

4.1 The benchmark

An overview of the crane winch is shown in Fig. 2(left), where the mechanical system is outlined with the most distinguished components, including the motor, the break system, the wrench drum. The motor is a three-phase induction motor. The motor is analyzed with vibration signals acquired with a tri-axial accelerometer positioned on the bearing housing at the gear side to monitor the transmission gearbox. In addition, a key phasor (8 events per cycle) is also installed on the motor shaft to get the motor rotor rotation speed.

The winch operation with its “up” and “down” modes is illustrated in Fig. 2(top-right). In order to acquire reliability information quickly, an accelerated life test is operated [5]. Two settings are used, a moderate load and a high speed applied to the winch for the first 20 months of operation, followed by a high load operation and a reduced speed, applied till the end of the test. See Fig. 2(bottom-right) for more details.

The signals and indicators are extracted as follows. From all the cycles, a large number of signals were discarded due to technical issues in the acquisition and low signal to noise ratio. The retained signals yielded 1141 samples for the “up” mode and 1138 for the “down” mode. For each sample, a set of 54 indicators were extracted, including indicators from the time domain (*e.g.* RMS, skewness, kurtosis and crest factor), from the frequency domain (*i.e.*, from appropriate frequency bands associated to the rotating parts), and from amplitude

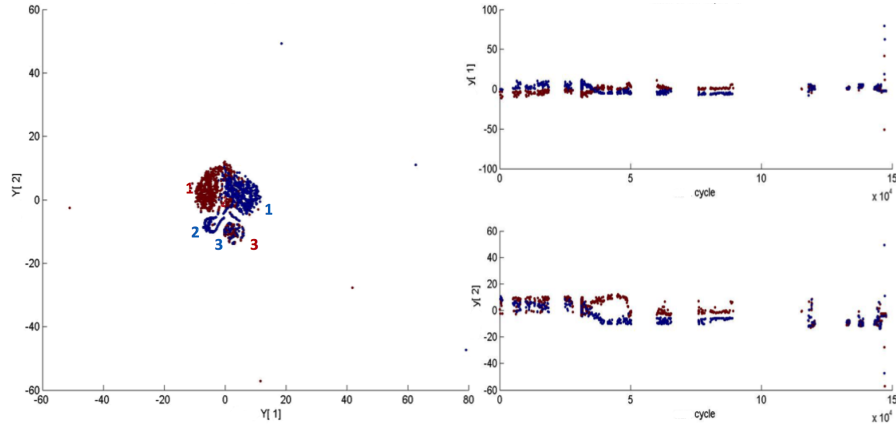


Fig. 3. Visualization in a two-dimensional space of the samples with the subset of relevant indicators (left panel), and the evolution in time of the first (top-right) and second (bottom-right) coordinates. Samples are shown from the “up” (in red \cdot) and “down” (in blue \cdot) modes, and the phases denoted by “1”, “2” and “3” for each mode.

and frequency modulation functions (investigates further the structure of the system under scrutiny). See [2] for more details. The indicators were computed from the stationary part given by the interval [3, 35] seconds for the “up” mode and [28, 50] seconds for the “down” mode. In the following, two different settings are confronted, one using all 54 indicators and one using a subset of “pertinent” indicators as recommended in [2].

4.2 Information retrieval and visualization results

The algorithm described in Section 2 was applied on the training samples acquired up to cycle #25000, where the estimated parameters set to $\lambda = 0.1$ and $\sigma = 40$ obtained from a grid search over a set of candidate values $\{10, 20, 30, \dots\}$. Fig. 4(left panel) shows the obtained representation in two-dimensions. In order to understand the dynamics of the evolution of the samples, Fig. 4(top- and bottom-right) presents separately the first and second coordinates as a function of the cycle number. As depicted in the figure, one can recognize three phases of evolution of the system, in both “up” and “down” modes (represented with red and blue dots, respectively), associated to the state of the training data (phase “1”), a transitional phase (phase “2”) and the last phase related to the second setting of high-load/low-speed (phase “3”).

4.3 Detection results

This section studies the relevance of the proposed method for detecting abnormal evolution, essentially from phase “1” to “2” and within each phase. As described

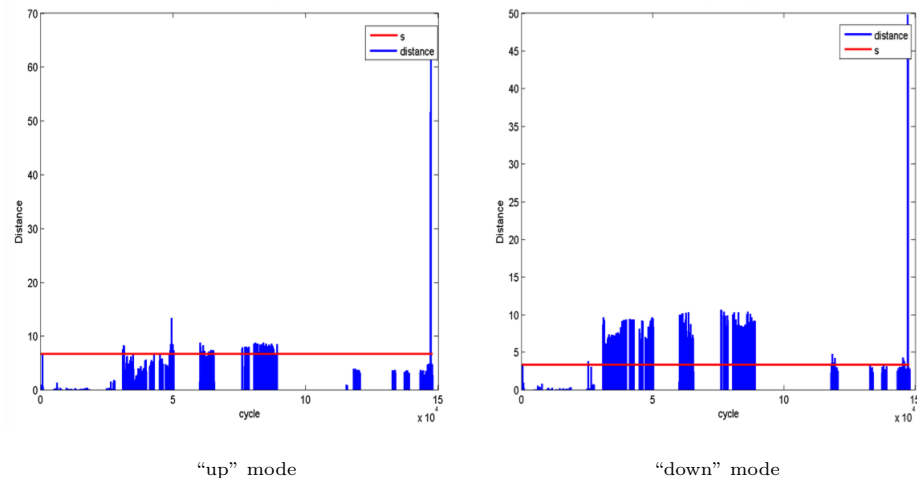


Fig. 4. Results of the detection method on the pertinent indicators, for the “up” (left panel) and “down” (right panel) modes, where s denotes the threshold and $distance$ is the measured minimum distance of each sample to the training samples.

in Section 3, the “up” and “down” modes are considered separately, each having a detector with its threshold. Investigating the subset of pertinent indicators, Fig. 4 shows the threshold s (in red) and the evolution of the distance rule for each sample (in blue), as given in (5). This rule allows to detect changes in regime between cycles #31139 and #31263 in the “up” mode, and between cycles #25375 and #31128 in the “down” mode. By using a sliding window to smooth the decision on three consecutive samples, one gets the range [#31126, #31128] for both modes. Fig. 4.3 shows the results obtained when all the indicators are used, showing larger variations at the last phase.

Finally, the performance of the proposed detection method is studied and compared to the state of the art. To this end, the well-known novelty detection with KPCA [7] is investigated. This method consists in defining a relevant subspace using the KPCA algorithm (a nonlinear variant of the PCA) and the novelty detection rule is based on the distance of the samples to this subspace. It is worth noting that this detection method has similar performance as one-class support vector machines and support vector data description [15]. To be comparable, a sliding window is used to smooth the decision as in the proposed method. Table 1 gives the range of first detected cycles for each method. It is easy to see that the proposed method detects the same cycles as KPCA, without the cumbersome computational complexity required for KPCA.

5 Conclusion and Future Work

This paper proposed the use of a nonlinear information retrieval method to represent data in a two-dimensional space, as well as representing the evolu-

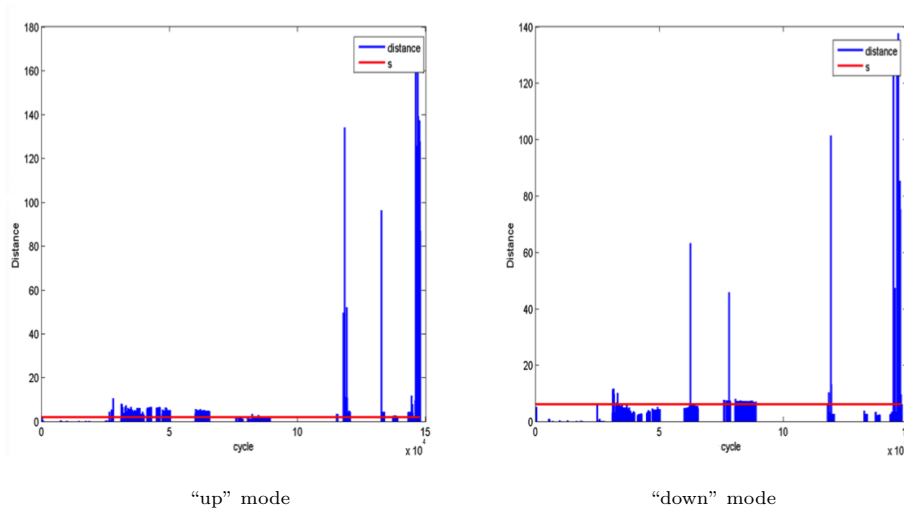


Fig. 5. Results of the detection method on all the indicators. Same legend as Fig. 4.

Table 1. Detection performance in terms of range for first detection cycles, when dealing with all indicators or only pertinent ones. The proposed detection method provides results that are comparable to the kernel principal component analysis (KPCA).

	Proposed method	Detection with KPCA
Pertinent indicators	[#31126, #31128]	[#31126, #31128]
All indicators	[#27375, #27750]	[#27375, #27625]

tion in time. Based on the extracted information, it described a novel detection method using distance and neighborhood information in order to monitor abnormal events. With its low computational complexity, the proposed method provided relevant results compared to the state-of-the-art detection method.

For future work, we will examine the relevance of these methods on supply current signals. We also want to revisit them in an online setting, namely having a new sample at each instant. To this end, two issues need to be addressed. On one hand, we will use the pre-image framework to overcome the out-of-sample issue in dimensionality reduction [11], as described in [10] for the conservation of similarity measures in both input and output spaces. On the other hand, an online sparsification criterion will be used to control the model order [9], in order to provide an online detection algorithm in the same spirit as [8].

Acknowledgments

This work is supported by the commission MLS (Manutention, Leverage et Stockage) of CETIM.

References

1. Assaad, B., Eltabach, M.: Vibration monitoring of winch epicyclic gearboxes using cyclostationarity and autoregressive signal model. In: G. Dalpiaz, et al. (eds.) *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*, Lecture Notes in Mechanical Engineering, pp. 261–271. Springer (2014)
2. Assaad, B., Eltabach, M., Antoni, J.: Vibration based condition monitoring of a multistage epicyclic gearbox in lifting cranes. *Mechanical Systems and Signal Processing* **42**(1–2), 351 – 367 (2014)
3. Cox, T.F., Cox, M.A.A.: *Multidimensional Scaling*, 2nd edition edn. Monographs on Statistics and Applied Probability. Chapman and Hall / CRC, London (2000)
4. El Hachemi Benbouzid, M.: A review of induction motors signature analysis as a medium for faults detection. *IEEE Transactions on Industrial Electronics* **47**(5), 984–993 (2000)
5. Escobar, L.A., Meeker, W.Q.: A review of accelerated test models. *Statistical Science* **21**(4), 552–577 (2006)
6. Hinton, G.E., Roweis, S.T.: Stochastic neighbor embedding. In: S. Becker, S. Thrun, K. Obermayer (eds.) *Advances in Neural Information Processing Systems 15*, pp. 857–864. MIT Press (2003). URL <http://papers.nips.cc/paper/2276-stochastic-neighbor-embedding.pdf>
7. Hoffmann, H.: Kernel pca for novelty detection. *Pattern Recognition* **40**(3), 863–874 (2007)
8. Honeine, P.: Online kernel principal component analysis: a reduced-order model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(9), 1814–1826 (2012)
9. Honeine, P.: Analyzing sparse dictionaries for online learning with kernels. *IEEE Transactions on Signal Processing* **63**(23), 6343–6353 (2015). URL <http://dx.doi.org/10.1109/TSP.2015.2457396>
10. Honeine, P., Richard, C.: A closed-form solution for the pre-image problem in kernel-based machines. *Journal of Signal Processing Systems* **65**(3), 289–299 (2011)
11. Honeine, P., Richard, C.: Preimage problem in kernel-based machine learning. *IEEE Signal Processing Magazine* **28**(2), 77–88 (2011)
12. Jolliffe, I.: *Principal Component Analysis*. Springer-Verlag, New York, NY, USA (1986)
13. Matic, D., Kulic, F., Pineda-Sanchez, M., Kamenko, I.: Support vector machine classifier for diagnosis in electrical machines: Application to broken bar. *Expert Systems with Applications* **39**(10), 8681 – 8689 (2012)
14. Moreira, M.L., Assaad, B., Eltabach, M.: New condition monitoring procedure for lifting cranes using motor current signal analysis. In: *International Conference Surveillance 7*. Institute of Technology of Chartres, France (2013)
15. Tax, D.M.J., Duin, R.P.W.: Support vector domain description. *Pattern Recognition Letters* **20**(11–13), 1191–1199 (1999)
16. Thorsen, O., Dalva, M.: Failure identification and analysis for high-voltage induction motors in the petrochemical industry. *IEEE Transactions on Industry Applications* **35**(4), 810–818 (1999). DOI 10.1109/28.777188
17. Vas, P.: *Parameter Estimation, Condition Monitoring, and Diagnosis of Electrical Machines*. Monographs in electrical and electronic engineering. Clarendon Press (1993)
18. Venna, J., Peltonen, J., Nybo, K., Aidos, H., Kaski, S.: Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research* **11**, 451–490 (2010)