

# TOWARDS ZERO-CONFIGURATION CONDITION MONITORING BASED ON DICTIONARY LEARNING

*Sergio Martin-del-Campo*<sup>\*†</sup>, *Fredrik Sandin*<sup>\*</sup>

<sup>\*</sup> EISLAB

Luleå University of Technology (LTU)  
971 87 Luleå, Sweden

<sup>†</sup> SKF-LTU University Technology Center

Luleå University of Technology (LTU)  
971 87 Luleå, Sweden

## ABSTRACT

Condition-based predictive maintenance can significantly improve overall equipment effectiveness provided that appropriate monitoring methods are used. Online condition monitoring systems are customized to each type of machine and need to be reconfigured when conditions change, which is costly and requires expert knowledge. Basic feature extraction methods limited to signal distribution functions and spectra are commonly used, making it difficult to automatically analyze and compare machine conditions. In this paper, we investigate the possibility to automate the condition monitoring process by continuously learning a dictionary of optimized shift-invariant feature vectors using a well-known sparse approximation method. We study how the feature vectors learned from a vibration signal evolve over time when a fault develops within a ball bearing of a rotating machine. We quantify the adaptation rate of learned features and find that this quantity changes significantly in the transitions between normal and faulty states of operation of the ball bearing.

**Index Terms**— Condition monitoring, feature extraction, dictionary learning, sparse representation, bearings

## 1. INTRODUCTION

Condition monitoring of machine elements is used to detect faults, reduce machine downtime and improve overall equipment effectiveness, for example by condition-based predictive maintenance. The requirements on the methods employed to achieve that go beyond fault detection, in particular in terms of prediction of faults [1] and detection of abnormal operational conditions. Early detection and characterization of emerging faults is a challenging problem because there are many variables that affect the operation of the machine and the characteristics of the fault. Maintenance operations rely on time and frequency domain features for diagnosis [1]. Expert knowledge is often needed to interpret the features and make decisions, which makes the process difficult to automate. Furthermore, condition monitoring methods are typically tuned to the application, the operating conditions and the type and location of the fault. Therefore, such methods are

expensive to maintain when machines have varying characteristics and evolve over time, for example as a consequence of maintenance and repair, which limits the scalability of the approach. Also, it is difficult to predict all failure modes. Similarly, approaches based on traditional pattern recognition methods require substantial amounts of labeled training data and the resulting methods are limited to the conditions for which the method was designed and trained [2].

Sparse representation of signals has attracted considerable interest in the last decade [3–6]. One type of sparse representation can be obtained by modeling signals as a linear superposition of noise and a small number of atomic waveforms (atoms) of particular shapes, amplitudes and shifts, so-called shift-invariant sparse coding [7, 8]. Using an approach known as dictionary learning the atoms can also be optimized to the signal [3, 6, 9], so that each particular atom represents structural features of the signal, which for example are excited by different physical processes. Such approximations are of increasing interest in signal processing with applications ranging from denoising, source coding, source separation, and signal acquisition. The problem of finding such sparse representations and optimal atoms is NP-hard in general. Therefore, suboptimal strategies based on convex relaxation, non-convex (often gradient based) local optimization or greedy search strategies are used in practise. Liu et al. [10] investigate the possibility that faults in a machine can be identified with multiclass linear discriminant analysis using dictionaries of atoms that are optimized to sets of signals corresponding to different fault conditions of a rotating machine.

In this paper we complement the study by Liu et al. by investigating how one dictionary of atoms changes over time in an online condition monitoring scenario. The dictionary is continuously optimized to a vibration signal, measured from a machine, that evolves from a normal state of operation to faulty conditions. We otherwise use a similar method for dictionary learning that is suited for online monitoring [11], and vibration signals from the same dataset [12]. The work presented here is novel because it focuses on online monitoring and the continuous evolution of an automatically learned dictionary, rather than supervised learning of multiple dictionaries for each fault condition. We demonstrate that devia-

tions from the normal state of the machine in principle can be detected via monitoring of the learned dictionary over time. We define an evolution rate for the atoms in a dictionary and demonstrate that this rate decreases to low values after some time of adaptation, and that it increases significantly when faults are introduced in the system. The resulting atoms are also useful for further classification and diagnosis of the condition [10, 11]. We find that some atoms characterize the vibration of the machine in both normal and abnormal operational conditions, while other waveforms are clearly associated with the faults. These preliminary results indicate that online monitoring of a learned dictionary is a potentially useful approach to zero-configuration fault detection. The approach also provides atoms representing inherent structural features in the signal that can be used for diagnosis and prediction.

## 2. SPARSE CODING AND DICTIONARY LEARNING

The model [11] used here was developed by Smith and Lewicki [13], and it is inspired by former work on sparse visual coding [14]. Smith and Lewicki discovered that atoms learned from speech data closely resemble cochlear impulse response functions (revcor filters), which indicates that speech is adapted to the ear [13]. Our working hypothesis is that features that characterize machines can be learned in a similar manner. The model decomposes a signal,  $x(t)$ , as a linear superposition of noise and atomic waveforms with compact support

$$x(t) = \sum_{m=1}^M \sum_{i=1}^{N_m} a_{m,i} \phi_m(t - \tau_{m,i}) + \epsilon(t). \quad (1)$$

The functions  $\phi_m(t)$  are *atoms* that represent morphological features of the signal and  $M$  indicates the total number of such atoms. The variable  $N_m$  refers to the number of instances of atom  $\phi_m$ , and the temporal position and amplitude of the  $i$ -th instance of atom  $\phi_m$  are denoted by  $\tau_{m,i}$  and  $a_{m,i}$ , respectively. The set of  $M$  atoms defines a dictionary

$$\Phi = \{\phi_1, \dots, \phi_M\}. \quad (2)$$

The values of  $\tau_i$  and  $a_i$  are determined with a matching pursuit (MP) algorithm [15, 16] with maximum a posteriori (MAP) optimization [17] for dictionary update. At each iteration  $n$ , MP is used to decompose the signal in the following steps:

1. Initialization:  $n = 0$ ,  $R_0(t) = x(t)$ ;
2. calculate cross-correlations between signal,  $x(t)$ , and all shift-invariant atoms,  $\phi_m(t)$ . The coefficients,  $a_{m,i}$ , takes the form

$$a_{m,i} = \langle R_n(t) | \phi_m(t - \tau_{m,i}) \rangle, \quad (3)$$

where the temporal position,  $\tau_{m,i}$ , is determined by the cross-correlation

$$\tau_{m,i} = \arg \max_{m,i} \langle R_n(t) | \phi_m(t - \tau_{m,i}) \rangle; \quad (4)$$

3. update of the residual

$$R_{n+1}(t) = R_n(t) - a_{m,i} \phi_m(t - \tau_{m,i}); \quad (5)$$

4. if the signal-to-residual ratio (SRR) or sparsity (number of samples over  $n$ ) reaches a predefined threshold the decomposition process stops;
5. each atom of dictionary,  $\Phi$ , is updated with a gradient procedure outlined below;
6. continue decomposition of the next set of (partially overlapping) signal samples.

The problem to learn the dictionary,  $\Phi$ , is the main challenge and opportunity of this approach, which makes it fundamentally different from traditional condition-monitoring approaches. The goal of this problem is to automatically calculate an optimal set of atomic waveforms,  $\phi_m$ , in the dictionary,  $\Phi$ , for a particular signal domain. The solution to this problem can be obtained by rewriting Eq. (1) in probabilistic form

$$p(x|\Phi) = \int p(x|a, \Phi) p(a) da \approx p(x|\hat{a}, \Phi) p(\hat{a}), \quad (6)$$

where  $\hat{a}$  is the maximum a posteriori (MAP) estimation of  $a$ ,

$$\hat{a} = \arg \max_a p(a|x, \Phi) = \arg \max_a p(x|a, \Phi) p(a), \quad (7)$$

that is generated by the MP [13]. The prior of the amplitude,  $p(a)$ , is defined to promote sparse coding in terms of statistically independent atoms [14] and it assumes that the likelihood,  $p(x|a, \Phi)$ , is Gaussian. This results in a learning algorithm that involves gradient ascent on the approximate log data probability [13]

$$\frac{\partial}{\partial \phi_m} \log(p(x|\Phi)) = \frac{1}{\sigma_\epsilon^2} \sum_i a_{m,i} [x - \hat{x}]_{\tau_{m,i}}. \quad (8)$$

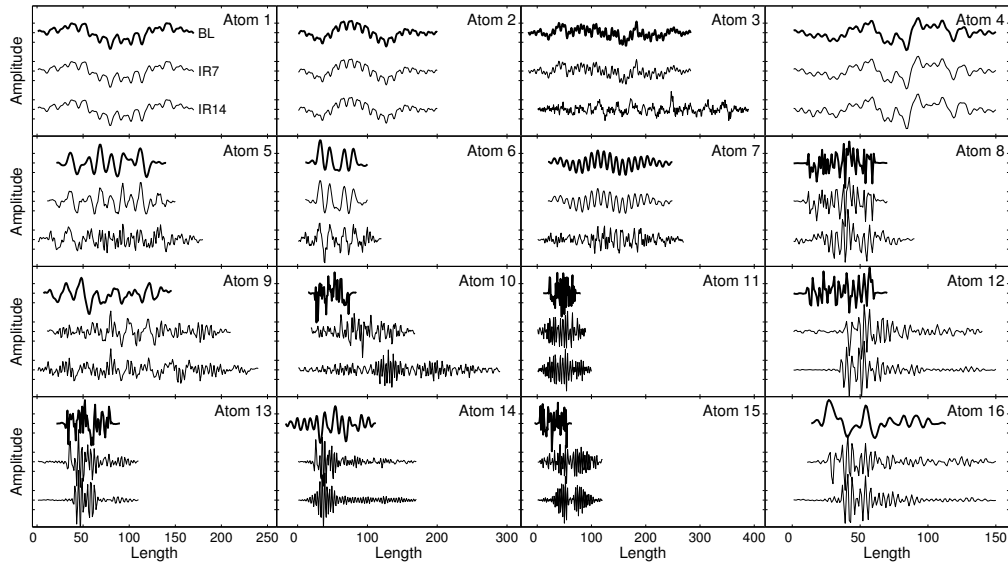
The gradient of each atom in the dictionary is proportional to the sum of residuals corresponding to the MP activation of that atom.

In order to use the gradient for optimization we introduce a *learning rate*, or stepsize parameter,  $\eta$ . Eq. (8) becomes

$$\frac{\partial}{\partial \phi_m} \log(p(x|\Phi)) = \frac{\eta}{\sigma_\epsilon^2} \sum_i a_{m,i} [x - \hat{x}]_{\tau_{m,i}}. \quad (9)$$

This means that the learning rate depends on the activation rate of atoms, which implies that the learning rate of atoms can be different, and that some atoms may not learn at all.

Several improvements of this methodology have been proposed, including methods to enforce orthogonality in the MP.



**Fig. 1.** Atoms learned from vibration signals corresponding to the BL, IR7 and IR14 cases, respectively. The atoms are ordered by ascending center frequencies in the IR14 case. All waveforms are normalized and have different lengths, see text.

Such methods improve the reconstruction accuracy significantly for noiseless signals, but the effect on denoising performance is moderate. Our method is comparable to that used by Liu et al. [10] and is motivated by the relatively low complexity and simplicity of the algorithm, which allows for online condition monitoring experiments in embedded systems.

We are interested in quantitative changes of the learned atoms resulting from changing conditions in a rotating machine. Skretting [18] proposes a dictionary distance measure as a means to quantify the similarity between two dictionaries. This approach is useful for diagnosis purposes but has limitations in an online monitoring scenario because only a subset of the atoms may change when a fault emerges, possibly resulting in high dictionary similarity. Therefore, we define the following *evolution rate* for each atom

$$\mu_a = 1 - \arg \max_b \text{crosscorr}(\phi_a(t), \phi_b(t - \delta)), \quad (10)$$

where  $\phi_a(t)$  is an atom of dictionary  $\Phi$  at time  $t$  and  $\phi_b(t - \delta)$  is the corresponding atom at a previous point in time,  $t - \delta$ . This quantity is calculated for each atom and it indicates how quickly individual atoms are changing. A value of zero means no change at all, while a value close to one means that an atom is uncorrelated with the most corresponding atom in the past.

### 3. CHARACTERIZATION OF ROTATING MACHINE WITH FAULT IN ROLLING ELEMENT BEARING

We apply the MP with dictionary learning approach to vibration data from a rotating machine at the bearing data center at Case Western Reserve University [12]. The vibration data was generated with a test rig consisting of an electric motor,

a torque transducer, a dynamometer and a ball bearing supporting the motor shaft. An accelerometer located at the drive end of the motor is used to record the vibration data. The accelerometer is sampled 12000 times per second. During data acquisition, the load varies between 0 HP and 3 HP, resulting in a varying motor speed from 1800 to 1730 rpm. We consider three different datasets in order to mimic the appearance and growth of a defect in the bearing, thereby simulating the evolution of the machine from a normal state of operation to a faulty state of operation. First, MP with dictionary learning is applied to 120 minutes of vibration data corresponding to a normal, non-faulty bearing. This is referred to as the baseline (BL) case and the resulting atoms are illustrated in Figure 1. Next, the atoms are further adapted to 120 minutes of data corresponding to a faulty bearing with a 7 mils (0.18 mm) diameter fault on the inner race. We refer to this as the IR7 case and the resulting atoms are also illustrated in Figure 1. Finally, the IR7 atoms are further adapted to 120 minutes of vibration data corresponding to a faulty bearing with a 14 mils (0.356 mm) fault on the inner race (IR14).

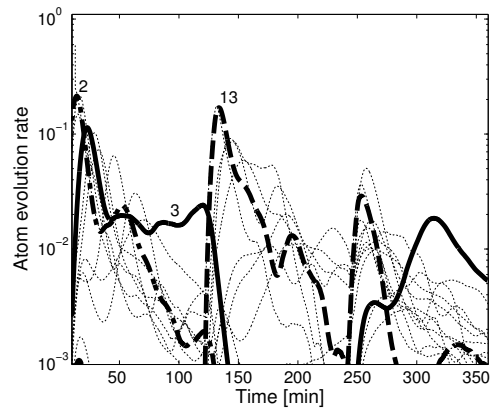
The vibration data is processed with our Matlab implementation of Smith and Lewicki's algorithm [13]. The dictionary initially contains sixteen atoms of length fifty sampled from a Gaussian distribution with zero mean. Dictionary learning is carried out using a signal window of 5 seconds duration (60000 samples). The windows are sampled randomly from the different load and rpm cases, thereby simulating a time-varying load on the rotating machine. Atoms are allowed to grow in length when the tail RMS exceeds a threshold [13] and are always normalized. MP is stopped at one order of magnitude reduction in the data rate, or at a 12 dB SRR.

| Atom # | Center freq. [kHz] |     |      | Event rate [s <sup>-1</sup> ] |     |      |
|--------|--------------------|-----|------|-------------------------------|-----|------|
|        | BL                 | IR7 | IR14 | BL                            | IR7 | IR14 |
| 1      | 0.1                | 0.1 | 0.1  | 55                            | 0   | 0    |
| 2      | 0.1                | 0.1 | 0.1  | 62                            | 0   | 0    |
| 3      | 1.1                | 1.1 | 0.4  | 99                            | 0   | 72   |
| 4      | 0.2                | 0.2 | 0.4  | 73                            | 0   | 3    |
| 5      | 1.0                | 0.9 | 0.5  | 66                            | 12  | 49   |
| 6      | 0.7                | 0.7 | 0.7  | 64                            | 3   | 27   |
| 7      | 1.0                | 1.0 | 1.1  | 50                            | 4   | 41   |
| 8      | 3.9                | 3.9 | 1.9  | 0                             | 3   | 63   |
| 9      | 0.7                | 1.5 | 1.9  | 58                            | 59  | 64   |
| 10     | 3.1                | 2.0 | 2.4  | 1                             | 116 | 70   |
| 11     | 4.4                | 3.2 | 3.1  | 0                             | 67  | 76   |
| 12     | 2.7                | 3.3 | 3.2  | 0                             | 100 | 74   |
| 13     | 2.4                | 3.5 | 3.2  | 0                             | 99  | 69   |
| 14     | 1.1                | 3.6 | 3.4  | 67                            | 108 | 69   |
| 15     | 2.8                | 3.1 | 3.5  | 0                             | 86  | 89   |
| 16     | 0.8                | 3.2 | 3.6  | 62                            | 107 | 82   |

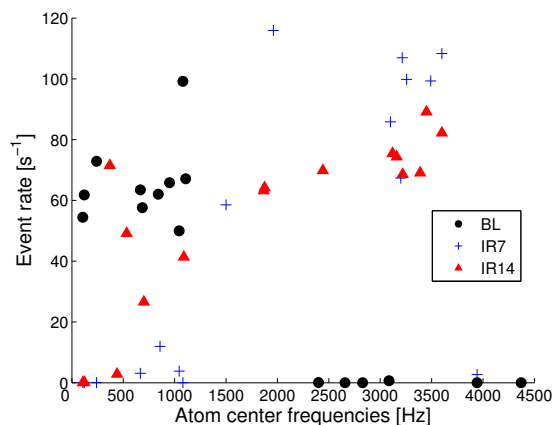
**Table 1.** Center frequencies and event rates of learned atoms.

The dictionaries resulting from the BL, IR7 and IR14 cases are shown in Figure 1, each including the sixteen atomic waveforms obtained at the end of a 120 minute adaptation time for each case. All waveforms are normalized and have the same scale. Each panel in Figure 1 illustrates one atom for the BL case (top), IR7 case (middle) and IR14 case (bottom). Atoms 1, 2 and 4 reach approximately stationary conditions after 120 minutes. Atoms 9, 10, 12, 13, 14, 15 and 16 change over time and enable distinction of the BL and IR7 cases. The difference between the IR7 and IR14 cases is evident from the time evolution of atoms 9, 10, 12 and 14. Furthermore, the differences between atoms 3, 5, 6, 7 and 8 distinguish the BL and IR14 cases.

Table 1 shows the center frequencies of the atoms in the three cases, calculated as the mean value of the power spectral density of each atom. By calculating the evolution rate (rate of change) of the atoms we notice changes in the characteristics of the rotating machine, which are associated with the introduction of a fault in the bearing. Figure 2 shows the evolution rate of all the atoms in the dictionary as defined by Eq. (10) and using  $\delta = 10$  minutes. Atom 3 stops evolving when the IR7 case is introduced after 120 minutes, this is represented by the disappearing bold line between 120 and 240 minutes, which is a consequence of the vanishing event rate, see Table 1. The center frequency of atom 3 is nearly identical in the BL and IR7 cases, see Table 1. Atom 3 continues to adapt after 240 minutes when the IR14 case is introduced. This is in agreement with Figure 1, which shows that atom 3 is similar for the BL and IR7 cases, while it has a different shape in the IR14 case. Atom 13 is inactive during the



**Fig. 2.** Evolution rate of atoms versus the time in minutes. The occurrences of the IR7 fault after 120 minutes, and the IR14 fault after 240 minutes affect the evolution rate of some atoms (bold lines) significantly, indicating that the characteristics of the machine changes at these points in time.



**Fig. 3.** Scatter plot of atom event rates versus center frequencies of atoms for the BL, IR7 and IR14 cases. The event rates are calculated during the last thirty minutes of vibration data in each case. The introduction of a fault in the bearing leads to learning and activation of atoms with high center frequency.

BL case, as indicated by the vanishing event rate in Table 1, but it starts to adapt in the IR7 case and eventually attains an impulse-like shape. In contrast, atom 2 adapts in the BL case and thereafter remains unchanged, see Figure 1. The center frequencies and event rates listed in Table 1, the evolution rate displayed in Figure 2 and the dictionary illustrated in Figure 1 provide complementary information about the three different operational conditions of the machine.

In Figure 3 we present a scatter plot of atom event rates versus the center frequency for the three cases listed in Table 1. It is evident that atoms with a lower center frequency occur in the BL case, while the cases including a bearing fault (IR7 and IR14) result in adaptation and activation of atoms with higher center frequencies. Furthermore, a comparison between the IR7 and IR14 cases reveals differences

in the event rates associated with some of the atoms. In summary, these results indicate that changes in the operational conditions and characteristics of a rotating machine can be automatically detected using unsupervised dictionary learning. Further work is required in the development of reliable measures for change detection during continuous monitoring of a rotating machine, including methods to avoid false positives associated with variations in the operation of the machine.

#### 4. DISCUSSION

We investigate the possibility to automatically characterize a rotating machine and detect when faults appear in the machine by monitoring a dictionary of learned atomic waveforms. We find that the shape, frequency and repetition characteristics of the atoms depend on the operational conditions of the machine considered here. Furthermore, we define the rate of change of atoms (the atom evolution rate) and illustrate that it can be useful for automatic detection of faults. These results motivate further experiments with more realistic failure modes and varying operational conditions. Further work is required to investigate and develop reliable measures for automatic change detection, possibly using a complementary knowledge base including atoms learned from similar machines with known operational conditions. In addition, deep learning extensions can be investigated for classification and prediction purposes. Dictionary learning offers a novel approach to online condition monitoring, which unlike most traditional techniques requires few assumptions about the machine and structure of the signal. Further work is needed to study the usefulness of the method under more realistic conditions such as speed, load and fault evolution. A method that requires a minimum of configuration is needed to enable scalable condition monitoring in the era of the Internet of Things

#### 5. ACKNOWLEDGMENTS

This work is supported by SKF, the Kempe Foundations, and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), grant IG2011-2025.

#### REFERENCES

- [1] H. Yang, J. Mathew, and L. Ma, "Intelligent diagnosis of rotating machinery faults - a review," in *Systems Integrity and Maintenance (ACSIM), 2002 Asia-Pacific Conference on*, September 2002.
- [2] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485 – 520, 2011.
- [3] M. Elad, "Sparse and redundant representation modeling – what next?," *Signal Processing Letters, IEEE*, vol. 19, no. 12, pp. 922–928, 2012.
- [4] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.
- [5] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1.
- [6] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 3rd edition, 2008.
- [7] R. B. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariance sparse coding for audio classification," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI2007)*, 2007.
- [8] B. Mailhe, R. Gribonval, F. Bimbot, and P. Vandergheynst, "A low complexity orthogonal matching pursuit for sparse signal approximation with shift-invariant dictionaries," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 3445–3448.
- [9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [10] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mech Syst Signal Pr*, vol. 25, no. 2, pp. 558 – 574, 2011.
- [11] S. Martin del Campo, K. Albertsson, J. Nilsson, J. Eliasson, and F. Sandin, "FPGA prototype of machine learning analog-to-feature converter for event-based succinct representation of signals," in *IEEE Int Workshop Mach Learn Signal Process*, UK, Sep 2013.
- [12] K.A. Loparo, "Bearing vibration data set," Case Western Reserve University, 2003, <http://csegroups.case.edu/bearingdatacenter/>.
- [13] E. C. Smith and M. S. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, no. 7079, pp. 978–982, 02 2006.
- [14] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [15] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397 –3415, dec 1993.
- [16] S. Krstulovic and R. Gribonval, "MPTK: Matching Pursuit made tractable," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2006, vol. 3, pp. 496–499.
- [17] E. Smith and M. S. Lewicki, "Efficient coding of time-relative structure using spikes," *Neural Computation*, vol. 17, no. 1, pp. 19–45, 2005.
- [18] K. Skretting and K. Engan, "Learned dictionaries for sparse image representation: properties and results," in *SPIE Proceedings*, 2011, vol. 8138.