RF-Ear: Contactless Multi-device Vibration Sensing and Identification Using COTS RFID

Panlong Yang, Yuanhao Feng, Jie Xiong, Ziyang Chen and Xiang-Yang Li
University of Science and Technology of China, China
University of Massachusetts Amherst, USA
Email: {fyhace, chenzi}@mail.ustc.edu.cn, {plyang, xiangyangli}@ustc.edu.cn, jxiong@cs.umass.edu

***Abstract*—Mechanical vibration sensing/monitoring plays a critical role in today's industrial Internet of Things (IoT) applications. Existing solutions usually involve directly attaching sensors to the target objects, which is invasive and may affect the operations of delicate devices. Non-invasive approaches such as video and laser methods have drawbacks in that, the former incurs poor performance in low light conditions, while the latter has difficulties to monitor multiple objects simultaneously.**

In this work, we design RF-Ear, a contactless vibration sensing system using Commercial off-the-shelf (COTS) RFID hardware. RF-Ear could accurately monitor the mechanical vibrations of multiple (up to 8) devices using a single tag: it can clearly tell which object is vibrating at what frequency without attaching tags on any device. RF-Ear can measure the vibration frequency up to 400Hz with a mean error rate of 0.2%. Our evaluation results show that RF-Ear can effectively detect 0.2cm screw loose with 90% accuracy. We further employ each device's unique vibration fingerprint to identify and differentiate devices of exactly the same model. We also show that RF-ear can monitor not just the vibrations but also a large range of mechanical motions. Comprehensive experiments conducted in a real power plant demonstrate the effectiveness of our system with outstanding performance.

***Index Terms*—RFID, Vibration Sensing, Identification, Contactless**

I. Introduction

Mechanical vibration sensing is a critical component of today's industrial Internet of Things (IoT) and plays a key role in machinery troubleshooting, working condition monitoring and industrial big data. Status quo solutions [1] are usually invasive, requiring attaching dedicated sensors to the targets. Non-invasive approaches such as video and laser monitoring [5], [24] are proposed for vibration sensing. However, the performance of video-based methods heavily depends on the lighting conditions and the sensing accuracy is usually low. It is also very challenging to capture subtle frequency variations with video. The laser method works in line-of-sight (LoS) and can only monitor one device at a time.

In the last few years, several RFID based approaches have been proposed including Tagbeat [30], Tagtwins [11], RED [34] and TagSound [21]. In these approaches, to measure the vibration frequency, RFID tags need to be mounted to the target device. These approaches have rather limited applicability because of the following issues. First, mounting tags to a machine often incurs extra labor cost and brings in safety issues. Second, although tags are usually small and light-weight, they could still affect the operation of a delicate device. We also notice that long-term vibration can easily loosen attached RFID tags. Thus, in this work, we target at monitoring the vibration in a contactless manner, without attaching any tag to the target device.

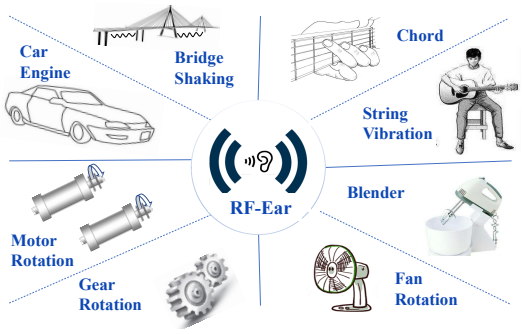


Fig. 1. Several candidate working scenarios.

As illustrated in Fig. 1, a desirable contactless RFID sensing system should be able to monitor different types of vibrations. It should also be able to concurrently monitor multiple objects and differentiate which object is vibrating at what frequency without attaching any tag on the objects. To realize this ambitious vision, we need to address several technical challenges:

- Using COTS RFID devices to monitor high vibrating frequencies (up to 400Hz) is challenging, as the COTS RFID reader only provides very low sampling rate, i.e., typically 40Hz at most [4].
- When multiple objects vibrate concurrently, the signal received at the RFID reader contains the mixture effect of multiple vibrating objects. It is challenging to separate the signals and extract the vibration information of each object [16].
- When multiple objects of same model are present, identifying which object is vibrating at what frequency is non-trivial. Objects of the same model have very similar vibration frequency and nearly the same outlook.

To address the first challenge, we propose a low-rank based signal recovery model to obtain the high frequency vibration signal from the low sampling rate signal. Specifically, we employ the modified OMP (Orthogonal Matching Pursuit) algorithm [22] to enable multi-frequency recovery from low sampling rate data. The original OMP algorithm suffers from high interference from the ambient noise, and we tackle this problem by employing a factor scaling scheme and an iteration scheme with voting. As a result, the resulting algorithm,

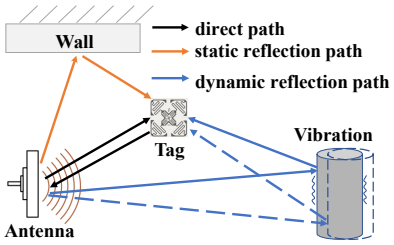


Fig. 2. Vibration Model

TABLE I
COMPARISON OF DIFFERENT VIBRATION SENSING SYSTEMS.

System	Error	RER	Contactless	LoS	Multi-obj	Anti-noise
accelerator [1]	0.1Hz	0.1%	No	Yes	No	Low
Camera [24]	1.0Hz	1%	Yes	Yes	Yes	Low
Laser [5]	0.01Hz	0.04%	Yes	Yes	No	Low
Tagbeat [30]	0.2Hz	0.3%	No	No	No	Low
Tagtwins [11]	0.2Hz	0.3%	No	No	No	High
RF-Ear	**0.1Hz**	**0.1%**	**Yes**	**No**	**Yes**	**High**

referred as MVOMP (Multi-Vibration Orthogonal Matching Pursuit), can deal with the noise much better when multiple objects are present.

To address the second challenge, we efficiently leverage the additive Gaussian noise to extract each individual vibration signal. We strike a balance fairly well between the accuracy and efficiency – by carefully setting the number of noise addition iterations. Such that, without knowing the number of devices, our system could select the interested signal from candidate frequencies at high accuracy, while differentiating from the noise efficiently.

To address the third challenge, we select 40 frequencies around the central vibrating frequency (the base frequency for a device) and use STFT (Short time Fourier Transformation) to construct the spectrogram as unique "fingerprints" for each device. We have conducted extensive experiments to show the effectiveness of the fingerprints and demonstrate its stability over a long period of time (e.g., several days) and across different locations.

As depicted in Table I, RF-Ear outperforms other systems in several important aspects. First of all, comparing with other non-invasive methods, our system is low-cost in supporting multi-objects monitoring. Even in NLoS cases, RF-Ear could still deliver high performance (details in Sec. VI). We would like to point out that the state-of-the-art solutions (such as [30]) still need to attach tags to the target for sensing and can only monitor one or two objects due to the low sampling rate. Last but not least, our system does not assume a prior knowledge of the exact number of devices for multi-target sensing.

In summary, our contributions are three-fold:

- We devise a COTS RFID system which is capable of monitoring the vibration of multiple targets simultaneously without attaching any tag on the targets.
- We employ a compressive reading approach to inspect high-frequency vibrations with low RFID sample rate. Our design could support monitoring vibration frequency up to 400Hz with low reader sample rate. The unique vibration fingerprint is successfully employed to identify each device at accuracy higher than 90% on average.
- We implement RF-Ear and evaluate its performance on commodity motors. Comprehensive experiments show that: for a single target, RF-Ear can inspect the vibration frequency at a mean error of $0.1Hz$; for simultaneously monitoring of 8 targets, the mean error is still as small as $1Hz$. For vibration amplitude, RF-Ear could effectively detect a 2mm change at 90% accuracy. We further demonstrate the applicability of our system on monitoring four different types of commonly seen equipments in a power plant, performing admirably well in all scenarios.

The remainder of this work is organized as follows. In Sec. II, we introduce preliminary knowledge on vibration model and low rank signal recovery method. In Sec. III, we outline the basic scheme and working flow of our proposed solution, and describe vibration sensing in Sec. IV. The data prepossessing model and fingerprint construction method are fully illustrated in Sec. V. Hereafter, performance evaluation is presented in Sec. VI. After reviewing the related work in Sec. VII, we conclude our work in Sec. VIII. The demo can be found at https://youtu.be/mAneBrHfut4.

II. PRELIMINARIES

A. Vibration Sensing Model

RFID Phase Model for Vibration Sensing: Typically in the RFID communication system, when there is a vibrating object nearby, the phase reading at the receiver is determined by signals of three paths, which are the *direct path* (Antenna → Tag → Antenna), *static reflection path* (Antenna → static non-target objects → Tag → antenna), and *dynamic reflection path* (Antenna → vibrating objects → Tag → antenna) as shown in Fig. 2.

In an RFID system with a vibrating object, the vibration will cause the length of the dynamic reflection path to slightly change and accordingly the signal phase to change. For RFID signal with a central frequency of 920MHz, a 5mm path length change will cause a phase change of 5.6 degrees. If we can capture the fine-grained phase change, we can thus monitor the signal path length change and accordingly sense the vibration.

B. Low-rank based Signal Recovery

The target device has a specific vibration frequency and thus the corresponding signal in frequency domain is sparse. The RFID protocol requires the tag to acknowledge the request from the RFID reader at a random time slot to avoid collision from multiple tags. The time domain signals collected at the RFID reader, denoted by y, can be converted into frequency domain signal S using FFT. Thus we have

$$y = \Phi s + \eta = \Phi \Psi^{-1} S + \eta, \tag{1}$$

where Φ is the measurement matrix, Ψ is the Fourier basis, and η is the measurement noise. The signal measured with

noise could be reconstructed reliably through solving an optimization problem according to Eq. (1)

$$\min \|y - \Phi\Psi^{-1}S\|_2^2, \quad \text{s.t.} \quad \|S\|_0 \leq K, \tag{2}$$

where K is the sparsity of S, which is also the number of different vibration frequencies. This optimization problem can be solved by Orthogonal Matching Pursuit (OMP) [22] method due to its simplicity and high efficiency.

The main idea of the OMP algorithm is to iteratively generate a list of atom indices $\mathbf{F}(S)$ and weighting scalars $\mathbf{P}(S)$ which represent a solution to the sparse signal representation. Particularly, $\mathbf{F}(S)$ and $\mathbf{P}(S)$ are defined as follows

$$\begin{aligned} \mathbf{F}(S) &= \{f_1, f_2, ...f_\tau \mid \forall i, S[f_i] \neq 0\}, \\ \mathbf{P}(S) &= \{p_1, p_2, ...p_\tau \mid \forall i, p_i = S[f_i]\}, \end{aligned} \tag{3}$$

where τ denotes the number of items. In each round, the algorithm iteratively adjusts the $\mathbf{P}(S)$ vector in a greedy manner, such that, the approximation error $R_i = \|y - \Phi\Psi^{-1}S_i\|_2^2$ can be reduced to the minimum, where S_i stands for the signal in the i^{th} iteration. Detailed algorithm description could be found in reference [22].

III. System Overview

Our system RF-Ear consists of two main parts. The first part is the *multi-vibration inspecting* module presented in Sec. IV. In order to deal with insufficient sampling rate, it needs to recover high frequency data from low sampling rate signals, and tame the noise to obtain target vibration frequencies. The second part is the *Vibrating-device Identification* module presented in Sec. V, identifying which object is vibrating at what frequency. In the rest of this section, we outline the design of RF-Ear as depicted in Fig. 3.

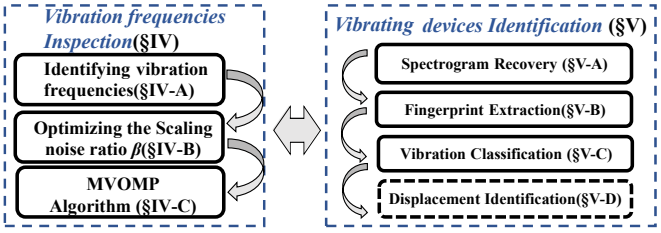


Fig. 3. System overview of RF-Ear, including vibration frequency inspection, and vibrating device identification.

Vibration Frequency Inspection: We start with a low-rank signal recovery technique to tackle the challenges caused by the low sampling rate of the RFID reader to enable high-frequency vibration sensing. We convert it to a minimum of l_0 norm problem and propose a modified version of Orthogonal Matching Pursuit (OMP) method to work with multi-object scenarios. Unfortunately, in many cases, we found that some of the noise frequencies have *very large* amplitude and can be regarded wrongly as vibration frequencies in frequency domain. To solve this problem, we carefully add a certain amount of white Gaussian noise to the received signal while keeping the overall amplitude unchanged through normalization. The vibration frequencies of objects can then be identified as those frequencies that have obviously decreased amplitudes after adding white Gaussian noise. We discuss the scaling factor optimization in §IV-B and the MVOMP algorithm in §IV-C.

Vibrating Devices Identification: Although frequency inspection could obtain the frequencies accurately for multiple vibrating objects, it could not support accurate identification among multiple devices. In order to identify which device is vibrating at what frequency, we need to go through another stage to realize this function. First of all, we recover the spectrogram for every vibration device. Next, we establish each device's vibration fingerprint using spectrogram, which consists of the vibration features in both the spatio-temporal domain and the frequency domain. We show the spatio-temporal *stability* of these fingerprints, as well as their *uniqueness* in §V-B. Hereafter, we leverage the CNN [18], [20] method and classify each device using the unique fingerprint.

IV. Frequency Inspection

A. Identifying vibration frequencies

In frequency domain, if the ambient noise is high, it may bring in frequency components that look similar to those caused by vibration. In order to discriminate between these two, our idea is to add *artificial noise* to *strengthen the noise amplitude but weaken the amplitude of actual vibration frequencies*. Consequently, we can identify those frequencies whose amplitude values decrease as *target* vibration frequencies of objects. As a matter of fact, we do not assume we have a prior knowledge about the exact number of devices, which differs from the previous work such as TagBeat [30]. In order to identify all the vibration frequencies without knowing how many devices are there, we take an iterative approach here. After each iteration of noise addition, we take a close look at the resulting SNR and identify the frequency components that become weaker. We end the iteration process when there is no significantly weaker component after noise addition.

Below we explain our noise addition algorithm in math model. Specifically, the background noise is regarded as Additive White Gaussian noise (AWGN). If we carefully tune the parameters and add an artificial AWGN to the frequency-domain signal S, the amplitude of the target vibration signal will decrease while the background noise will increase after normalization.

We define the received signal as y. If the vibration signals are s_v and the noise is η_0, $y = S_v + \eta_0$. Assume the SNR of the received signal y is α, thus, $\mathbf{E}(S_v) = \alpha\mathbf{E}(\eta_0)$. Here $\mathbf{E}(\cdot)$ represents the power of signals. Next, we add an AWGN η_1 whose amplitude is $1/\beta$ of the original signal y, *i.e.*, $\mathbf{E}(\eta_1) = \frac{1}{\beta}\mathbf{E}(y)$. Then $\mathbf{E}(\eta_1) = \dfrac{\alpha+1}{\beta}\mathbf{E}(\eta_0)$. After we add AWGN noise to y, we normalize the new signal to keep the total power unchanged. We define the new signal as y', the total power of noise in y' as $\mathbf{E}(\eta')$, and the vibration signal in y'

as s'_v. Then we have

$$\frac{\mathbf{E}(S'_v)}{\mathbf{E}(y')} = \frac{\mathbf{E}(S_v)}{\mathbf{E}(S_v)+\mathbf{E}(\eta_0)+\mathbf{E}(\eta_1)} = \frac{\alpha\beta}{(\alpha+1)(\beta+1)},$$
$$\frac{\mathbf{E}(\eta')}{\mathbf{E}(y')} = \frac{\mathbf{E}(\eta_0)+\mathbf{E}(\eta_1)}{\mathbf{E}(S_v)+\mathbf{E}(\eta_0)+\mathbf{E}(\eta_1)} = \frac{\alpha+\beta+1}{(\alpha+1)(\beta+1)}.$$

We also have $\mathbf{E}(y) = \mathbf{E}(y')$ then

$$\frac{\mathbf{E}(S'_v)}{\mathbf{E}(S_v)} = \frac{\mathbf{E}(S'_v)}{\mathbf{E}(y')}\frac{\mathbf{E}(y)}{\mathbf{E}(S_v)} = \frac{\beta}{\beta+1} < 1, \tag{4}$$
$$\frac{\mathbf{E}(\eta')}{\mathbf{E}(\eta_0)} = \frac{\mathbf{E}(\eta')}{\mathbf{E}(y')}\frac{\mathbf{E}(y)}{\mathbf{E}(\eta_0)} = \frac{\alpha+\beta+1}{\beta+1} > 1. \tag{5}$$

Algorithm 1: MVOMP algorithm

Input: signal y, FFT base Ψ, No. Possible Items τ
Output: Recovered vibration-induced signal S

1 $\beta \leftarrow 0.5$; $Suitbeta \leftarrow false$; $n \leftarrow 100$;
2 $S_0 \leftarrow OMP(y, \Psi, \tau)$ using previous method OMP.;
3 $F_0 \leftarrow F(S_0)$ to recover τ frequencies.;
4 $P_0 \leftarrow P(S_0)$ are the corresponding amplitude values;
5 **while** $Suitbeta = false$ **do**
6 **for** $i = 1 \rightarrow n$ **do**
7 $y_i \leftarrow Addnoise(y, \beta)$; $S_i \leftarrow OMP(y_i, \Psi, \tau)$;
8 Recover all frequencies and amplitude values as $F_i \leftarrow F(S_i)$; $P_i \leftarrow P(S_i)$;
9 **end**
10 **if** $\forall 1 \le i \le n, F_i = F_0$ **then**
11 $Suitbeta \leftarrow true$;
12 **end**
13 Set $\beta = \beta + 0.5$
14 **end**
15 **for** $i = 1 \rightarrow \tau$ **do**
16 $Vote \leftarrow |\{P_j[i] < P_0[i] \mid \forall j, 1 \le j \le n\}|$;
17 **if** $Vote > 2/3 \times n$ **then**
18 $S[F_0[i]] = P_0[i]$;
19 **end**
20 **end**

Thus, after adding AWGN noise and normalization, the power of the vibration signal is decreased, and the noise power is increased. Furthermore, the magnitude of the increasing and decreasing can be controlled by the parameter β.

B. Optimizing the Scaling noise ratio β

There are two issues when we apply the above method. The first issue is searching the optimal β^*. The noise added to the original signal should not be *too large* (noise dominated, and difficult to find the interested signal) or *too small* (no significant difference after adding noise). Specifically, we establish a frequency set for the result of OMP algorithm and select a big enough threshold as its size (how many frequencies are there in the set) to ensure that all the vibration frequencies are included. The threshold can be set via our experiment presented in §VI-B. Then we test artificial AWGN with different β (from small to large at a step size of 0.5) and compare the frequencies before and after adding noise. If β is too small, the noise will be large enough to affect the frequency set (there will be new frequencies appear in the frequency set). Otherwise, if β is too large, the added AWGN noise is very small and thus the change of the amplitude is not significant enough to be employed to identify the vibration frequencies from noise. Assume the OMP output of the original signal y is S, and the OMP output of the new signal y' is $S'(\beta)$. The optimal value of β is β_0, then

$$\forall \beta < \beta_0, \quad \mathbf{F}(S'(\beta)) \neq \mathbf{F}(S),$$
$$\forall \beta \ge \beta_0, \quad \mathbf{F}(S'(\beta)) = \mathbf{F}(S). \tag{6}$$

We set the initial β as 0.5 and increase β gradually until we find out the first β which satisfies $\mathbf{F}(S(\beta)) = \mathbf{F}(S)$.

The second issue is removing the negative impact of randomness of the AWGN added. We add AWGN with the optimal β^* to the original signal and perform normalization operation. Ideally, the white noise will increase and the vibration frequency components will decrease. However, in reality, the added noise may also constructively combine with the vibration frequency components although the chance is small.

To remove this small uncertainty brought in, we repeat this process of adding noise n times and vote to identify the vibration frequencies. The optimal vote rate is set according our experiment in §VI-B.

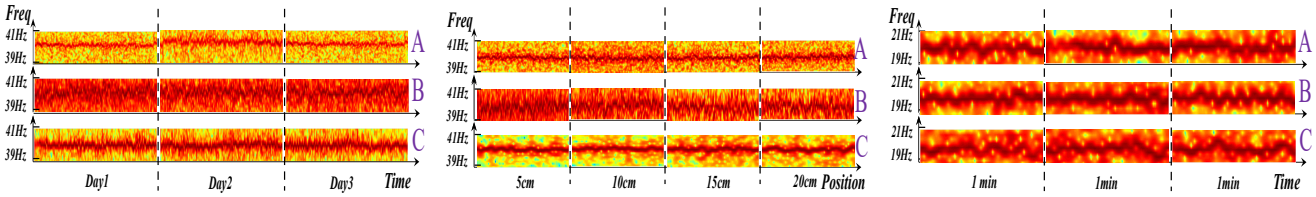
C. MVOMP: Putting It All Together

As aforementioned factor β is set, we now have a modified OMP algorithm, enhanced by the adaptive noise addition method shown in Algorithm 1. We call it Multiple Vibrations Recovery-based OMP (MVOMP). From Line 1 to 4, we set the initial values for processing, and from Line 4 to 14, we search for the most suitable β according to the OMP feedback before adding noise. The value for β is increased at a small step size of 0.5 to ensure that the found β^* will be close to the optimal. From Line 15 to 20, we compare and extract the vibration frequencies from the candidate items with a simple voting scheme. Note that the number of items (frequencies) τ is not exactly known. Thus we first employ some possible number of items in our algorithm, evaluate the performance and select an optimized value, which is depicted in Sec. VI-B.

V. Device Identification

A. Spectrogram Recovery

Recovery on Multiple Frequencies: Most real-world periodic signals are not exactly so and there are actually small changes in the period of the signal over time. We refer such a broader class of signals as *quasi-periodic* [25]. Variation in base-frequency f of a quasi-periodic signal is reflected in its Fourier transform, which contains non-negligible amplitudes in a small band around f. The essence of MVOMP algorithm is to search the frequency point with the highest amplitude iteratively. After removing the noise frequencies, the results are the K vibration frequency components, each associated with one amplitude representing the power.



(a) Temporal stability (b) Location stability (c) Non-distinguishable with short duration

Fig. 4. Fingerprints for three same model devices on different locations over several days

On Parameter Setting: In our system RF-Ear, we choose X frequency points that are equally spaced around the base-frequency with an interval size of 0.05Hz between adjacent frequency points. These frequency points will be used by OMP algorithm to search their amplitudes. As a result, when given K vibrating devices with potentially K different base-frequencies (one for each device), there are a total of $(X+1)\times K$ frequencies and amplitudes. Note that from our extensive experiments, the frequency variation range is up-to 2Hz, so we set X as 40 in our experiments to cover the 2HZ variation range. Note that when the distances between the vibration device, the RFID reader, and the RFID tags vary, the amplitude of the signal vary. To remove the negative impact of such difference in amplitude values, a normalization of amplitude is implemented to ensure all the vibration signals have similar amplitudes.

B. Device Fingerprint Extraction

Basic Method: To achieve the objective of telling which device is vibrating at what frequency, another challenge is to differentiate between similar devices such as devices of the same model. Our device identification scheme is based on a key observation that even devices of the same model can still exhibit dramatically different frequency-amplitude patterns over time. It can be clearly visualized in the time-frequency plot. The longer the time period, the more obvious this difference is. To distinguish the patterns, we apply the well-studied CNN [18], [20] method. The input to the CNN is the time-frequency information. For Short Time Fourier Transform (STFT), the window-size is set as 2s and the overlap segment is set as 0.2s. The result of STFT is transformed and saved as a heat-map, which is shown in Fig. 4.

Spatio-temporal Stability: Before applying these features for device identification, we need to test the *stability* and the *resolution* of the proposed fingerprints. To evaluate the feasibility and stability of our proposed fingerprint features, we conduct extensive tests with three devices at four different locations and collect fingerprints over 30 days. As illustrated in Fig. 4a and 4b, the fingerprints at different locations are presented over a period of time. It exhibits clear differences among devices. A, B and C are three different machines of the same model. As illustrated in Fig. 4a, we can see that the fingerprints show excellent stability over a long time period. From Fig. 4b, we can see that even when the target is slightly moved for a few centimeters, the pattern is still relatively stable.

Feasibility of Fingerprints: One may wonder that why similar devices (same model) exhibit significant amount of differences on the spectrogram? An inherent reason lies in the high frequency vibration and long period of data collection, which provide rich information on vibrating objects. As a counter example, if we lower the vibrating frequency and decrease the data collecting period, the fingerprints among different devices are hardly discernible, as shown in Fig. 4c.

C. Vibration Classification

Based on the fingerprint features extracted from different devices, we differentiate devices with three classifiers: KNN (K-Nearest Neighbor), SVM (Support Vector Machine) and CNN (Convolutional Neural Network). Among these three commonly used classifiers, CNN achieves the best performance. We use the hybrid CNN structure [10] which can achieve high resolution and the complexity is still relatively low. Specifically, three convolution layers and three pooling layers are used, followed by two fully connected layers and one output layer. The sizes of convolution kernels are 11 x 11, 5 x 5 and 3 x 3 for each layer respectively, and the pooling size is 3 x 3, with the stride of 2. We first collect the features of the devices when they vibrate alone as training set, and collect the features of the devices when they vibrate concurrently as test set. Moreover, we intentionally add some new data sets to test the identification accuracy. The performance of these three classifiers is presented in Table II. CNN achieves the best performance because CNN can capture the deep level features of the sample and does not rely on manually defined features.

TABLE II
PERFORMANCE OF DIFFERENT CLASSIFIERS.

Classifier	Identification Accuracy	Time Cost
KNN	90%	1500ms
SVM	89%	1000ms
CNN	94%	700ms

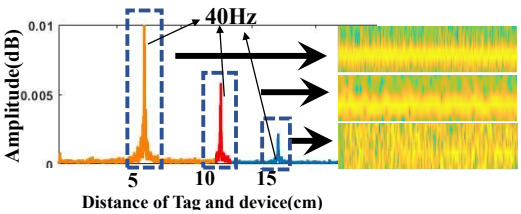


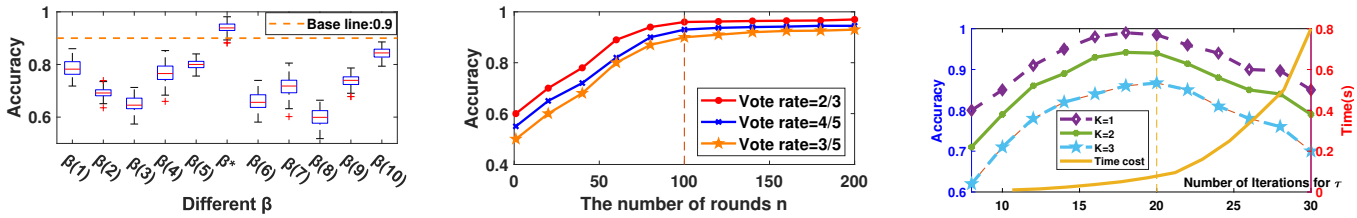
Fig. 5. Fingerprint without normalization

D. Displacement Identification

When machines vibrate at a high frequency (e.g., $\geq 50Hz$), it would possibly lead to device displacement due to the imbalanced mechanical vibration or poor fixtures. Although this kind of displacement may be discernible, it is vital for

(a) Experimental setup (b) Different blocking scenarios (c) guitar, blender,fan (d) four equipments

Fig. 6. (a) Experimental setup, (b) NLOS evaluation, (c) Three typical applications, and (d) Power plant.



(a) Impact of noise ratio β (b) Impact of iterative rounds n (c) Impact of number of items τ

Fig. 7. Illustration of optimizing parameters for iterative rounds n, noise ratio β, and number of items τ.

device monitoring in some safety related production. Leveraging aforementioned classifier scheme, we can also detect this kind of small displacement at high accuracies. A small 2cm displacement can be detected 90% of the time. We will report the detail evaluation results in Sec. VI-D.

One may wonder that, does this result conflict with the aforementioned device fingerprint feature stability in spatial domain? Actually, it does not. The reason is as follows. When we recover the spectrogram, we normalize the amplitude to remove the impact of device position. While here for displacement identification, we do not perform this normalization step. Thus, the training data as well as the device fingerprints consist of the displacement information. As shown in Fig. 5, when there is no normalization for the device fingerprints at different distances, it could be relatively easy to see the difference brought in by different displacements.

VI. Evaluation Results

A. Experiment Setups

Hardware Setup: We use an ImpinJ Speedway R420 reader [4] which adopts the EPC Gen2 standard [3]. Four types of RFID tags from Alien [2], with sizes of $2.0cm \times 2.0cm$, $7.5cm \times 2.0cm$, $10cm \times 2.5cm$, and $10cm \times 5cm$ are used in our experiments. RF-Ear works in the 920-926 MHz band and a directional antenna with $8dB$ gain in used. For the purpose of eliminating the impact of network latency, we adopt the timestamp scheme provided by the reader instead of the received time.

Software Setup: We adopt the Impinj LLRP Tool Kit [6] to communicate with the reader. ImpinJ reader improves this protocol to support the phase reading report. The client software is implemented using C# for network connection and Matlab for signal processing. In our experiment, we run our system at a Lenovo PC, equipped with 2.5 GHz Intel Core i5 and 8G memory.

Basic Working Scenario: As shown in Fig. 6a, there are two motors on the table, separated by a distance of $5cm$. Their vibration frequencies can be tuned from 0 to $6,000$ RPM (Revolution Per Minute) with a speed controller. A *single* RFID tag is placed near to the motors (*i.e.*, 5cm apart from each motor), and the distance to the reader antenna is set as $1.5m$ by default. Fig. 6b depicts the NLoS setting, and For application case study, we test our system in the lab with guitar, blender and fan in Fig. 6c. We further evaluate the performance of our system in the wild in a real power plant, which is shown in Fig. 6d. We use a COTS laser tachometer to acquire the ground truths of the frequency measurements.

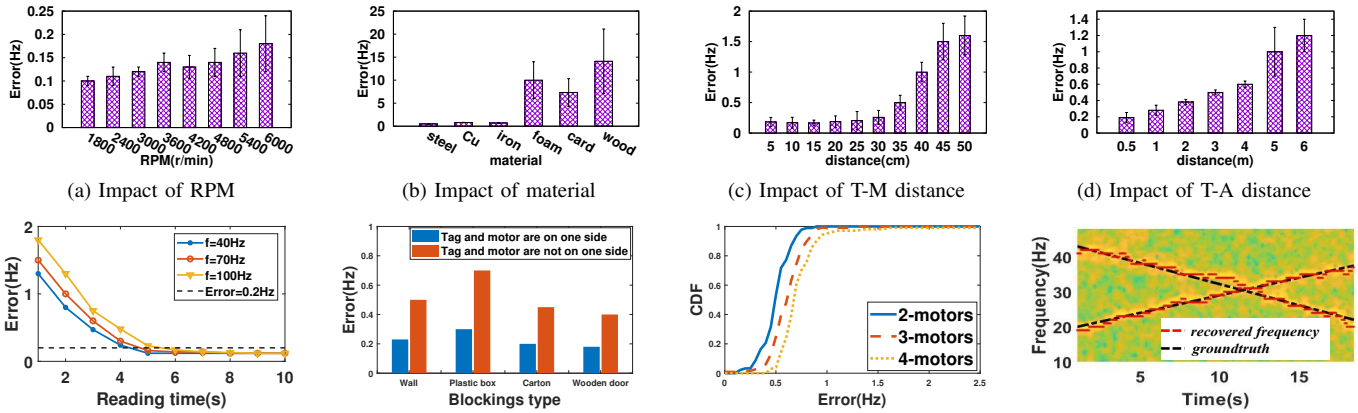
B. Parameter Setting for MVOMP

We explain how we set the three important parameters for the MVOMP algorithm. The first one is the scaling noise ratio β, which is important for noise addition scheme, and the second one is the number of iterative rounds n, which is the key parameter for noise taming process. The last parameter is τ, which affects the algorithm execution speed and performance.

The scaling noise ratio β: To show the effectiveness of this parameter, we employ ten random β values, denoted as $\{\beta_1, \beta_2, \cdots, \beta_{10}\}$. For each of these values, we run our tests n times. The performance is shown in Fig. 7a and we can see that the frequency sensing accuracy is highest when $\beta = \beta^*$.

The number of rounds n: As shown in Fig. 7b, the accuracy of vibration extraction increases with n in all three curves. However, when n is greater than 100, the performance almost saturates. Thus, we set the number of rounds n to 100. We also tested the effect of different voting rates. As shown in Fig. 7b, as long as the number n is larger than 100, all the 3 voting rates achieve reasonably good performance. We also test other voting rates and the accuracy is stable above 90% when n is set more than 100.

Number of Items τ: The number of items (frequencies) τ is highly related to the sparsity of signal S. Value τ should be



(a) Impact of RPM (b) Impact of material (c) Impact of T-M distance (d) Impact of T-A distance

(e) Impact of reading time (f) Impact of blocking (g) Impact of motor number (h) Frequency changes with time

Fig. 8. Performance for Vibrations Sensing under the Impact of Different Factors.

set larger than the number (K) of signals. However, τ should not be set "too large", because it would incorporate more noise components, which would lead to decreased accuracy and also higher computational cost.

Thus, we consider two factors for setting the value τ. The first is the number of vibrating devices, and the second factor is the time cost. Consider a scenario with 8 devices as depicted in Fig. 7c, the accuracy of vibration extraction is highest when the value of τ reaches 18. The accuracy drops after τ values is further increased. Furthermore, as depicted in Fig. 7c, the computational cost is larger when τ increases. As a result, we set τ as 18 for 8 concurrent vibrating devices. Note that although the parameters chosen for 8 vibrating devices may not be optimal for 4-6 vibrating devices, our experiment results show that these parameters still achieve reasonably good performance.

C. Results on Frequency Inspection

We first evaluate the performance of single-object vibration sensing.

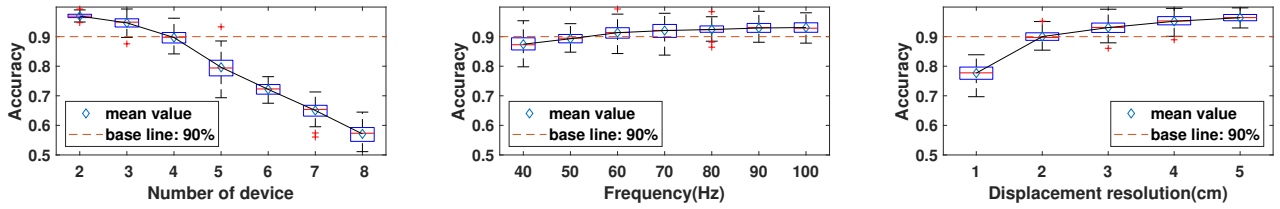
Impact of RPM: To investigate the impact of spinning speed on RF-Ear, we increase the RPM from $1,800$ to $6,000$ at a step size of 600. Each experiment is repeated 50 times. The average results as well as the standard deviations are presented in. Fig. 8a. We find two inspiring observations:

- The overall mean error is as small as $0.14Hz$ and the average relative error rate is 0.14% (*i.e.*, the ratio between the error and the true base frequency). Compared with Tagbeat [30] and Tagtwins [11], we achieve even better performance without a need of attaching tags to the targets.
- For a speed of $6000RPM$, the mean error is $0.18Hz$, which is slightly higher than lower speeds. This is because the motor shakes stronger at higher frequency, which introduces more noise.

Impact of target Material: Next, we study the influence of target materials. In reality, the vibration devices could be made of different materials, which would impose significant diversity on the strength of the reflected signals. We choose three types of metallic materials (*i.e.*, steel, copper, iron) and three types of nonmetallic materials (*i.e.*, foam, cardboard, wood) for our evaluation. Keeping the motor spinning speed at $2,400$ RPM (*i.e.*, 40Hz), we repeat the experiments 50 times for each material. Fig. 8b plots the mean errors on frequency estimation. It can be seen that the mean error for metallic materials is 0.6Hz while that for nonmetallic material is around 10Hz. The much larger error for nonmetallic materials is mainly due to the much weaker signals reflected from the nonmetallic materials.

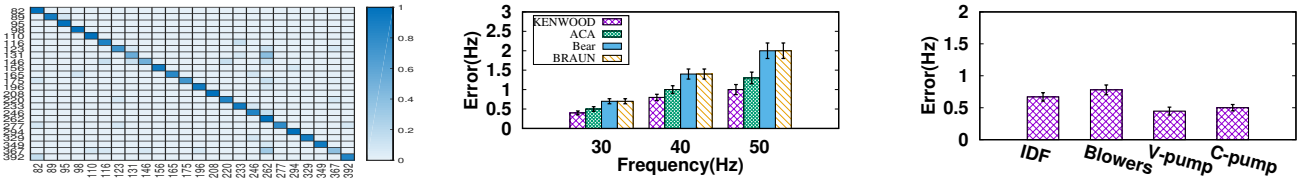
Impact of Distance:

- Distance between Tag and Objects: We vary the distances from $5cm$ to $50cm$ at a step size of 5cm. We keep the vibrating frequency as 40Hz and the distance between tag and antenna as $0.5m$. We repeat experiments 50 times at each distance. Fig. 8c shows the mean errors and the standard deviations. We observe that when the distance is smaller than $35cm$, all mean errors are below $0.5Hz$. The mean errors and standard deviations increase significantly when the distance is larger than $35cm$. This implies the performance of RF-Ear does get affected when the distance between tag and vibration source becomes too large. This is mainly because our sensing is contactless, thus relying on the weak reflected signal. The reflected signal becomes too weak to be employed for accurate sensing when the distance becomes larger than 35cm. To increase this sensing range, one possible way is to employ more efficient directional antenna which can focus energy at one direction. However, this will decrease the sensing range at other directions a trade-off exists.
- Distance between Tag and Antenna: The communication range between the tag and antenna of our COTS RFID hardware is around 6 meters. Thus we vary the distance from 0.5m to 6m and repeat the experiments 50 times at each distance. The distance between the tag and motor is fixed as 5cm. Fig. 8d shows the result. The mean error increases when we increase the distance. This is expected because the received signal becomes weaker with longer signal propagation distance in the air. To achieve the best sensing performance, we suggest to keep the distance between the antenna and reader in the range of $0.5m \sim 4m$. Note that when the distance



(a) Impact of Device number (b) Impact of frequency (c) Displacement resolution

Fig. 9. Performance for Vibrations Identification under the impact of different factors.



(a) Confusion Matrix for 24 tones (b) Performance on Blenders (c) Performance on Fan

Fig. 10. Performance on different cases: playing guitar, a working blender, and four equipments in power plant.

between tag and antenna is smaller than $0.5m$, the sensing performance is still good. However, the size of the sensing area becomes smaller and thus the target is required to be very close to the tag.

Impact of Blocking: We now test our system in more challenging NLoS scenario, such as behind a wall. As shown in Fig. 6b, four different cases (one or two motors are put inside a carton, a plastic box, behind a concrete wall, and behind a wood door) are evaluated. The distance from the tag to each motor is 5cm while the distance to antenna is 1.5m. The results are shown in Fig. 8f. We can see that NLoS blockage does decrease the sensing accuracy. However, even with the blockage, the mean error is still within $0.4Hz$, which is good enough for a lot of applications. These results demonstrate our system can still work in challenging NLoS scenarios.

Impact of Number of Motors: Fig. 8g depicts the CDF plot of the frequency error with various numbers of motors. In our experiment, we set the number of motors as 2, 3, 4 respectively and conduct 100 repeated experiments for each number. The frequencies of *motor1*, *motor2* and *motor3* are set as 30Hz, 40Hz, and 50Hz respectively. The distance from the tag to each motor is 5cm while the distance to antenna is 1.5m. It can be seen from Fig. 8g that the mean error is less than 0.6Hz for 50% of the measurements and less than 1Hz for 90% of measurements. We also observe that the mean error increases with the increasing number of motors. This is mainly due to higher interference between motors.

Performance with Varying Frequency: Lastly, we test RF-Ear in situations when the rotation frequency is not a constant but increases or decreases. We smoothly increase the rotating speed from 20Hz to 40Hz in 20s and then decrease it back to 20Hz in 20s. As shown in Fig. 8h, the obtained vibrating frequencies match the ground-truths very well and the mean error is as small as $0.1Hz$.

Other factors: We also test the impact of tag type and the deployment orientation of tag. We find that the tag size does affect the accuracy (the bigger, the better) while the tag orientation does not have significant influence on the accuracy.

D. Device Identification

In this subsection, we evaluate the system's capability in device identification. Note that, for device identification, we work on the most challenging scenario: the devices are of the same model and rotate at exactly the same frequency.

Impact of Number of Targets: We evaluate the device identification accuracy up-to 8 simultaneously vibrating devices. Fig. 9a depicts the identification accuracy with different number of devices. With more devices, it is expected to see the performance degrades. Note that for vibration sensing, even for 8 devices, we can still achieve a 90% identification accuracy. To achieve higher than 90% accuracy, the supported maximum device number is 4.

Impact of Frequency Diversity: Fig. 9b depicts the identification accuracy across different central vibration frequencies. With the increased central frequency, the identification accuracy slightly increases. This is because with higher central frequency, the spectrogram actually provides us more information and the difference between devices becomes clearer.

Impact of Screw Loose: We further evaluate if a small change such as screw loose at the motor can be detected. We employ a screw with a weight of 40 g and a length of 4.1 cm. The radius of the screws is around 3mm. We place the screw partially inside the motor and the amount of length inside the motor is decreased from $4.1cm$ (a screw loose of 0cm) to $3.1cm$ (a screw loose of 1cm) at a step size of $0.1cm$. The weight of the motor is around 2Kg, which is much larger than the screw. We employ the CNN algorithm to see if the small difference caused by screw loose can be detected. Our results show that when the screw loose is 0.2cm, it can be detected with an accuracy of 90%. When the screw loose is larger than 0.3cm, it can be detected 100% of the time.

Displacement Identification: The displacement identification ability is evaluated and shown in Fig. 9c, where a 2cm

location displacement could be identified at over 90% of accuracy.

E. Application Studies in the Lab

In this section, we show the usage of the proposed system with real-word applications. We apply our system to monitor the guitar play, blender vibration and fan rotation.

The experimental scenario is shown in Fig. 6c. The frequency of each guitar string is calibrated by the GuitarTuna [13]. The frequency is in the range of 80-400Hz in the first three fret. Compared with the motor vibration which lasts for a long period of time and is relatively stable, string vibration could attenuate quickly within $0.5 \sim 2s$, making the sensing more challenging.

Sensing Accuracy: After data preprocessing, we obtain the signal of each vibration and employ the proposed MVOMP algorithm to recover the vibration frequency of the string. We use the accuracy of tone to evaluate the performance because our goal here is to capture which tone the guitar is playing instead of the exact frequency. If the result is in the range of $(f-\delta, f+\delta)$, it can be regarded as correct, where the ground truth is f and the δ is the threshold. In our experiment, we set the threshold δ as 1Hz. Fig. 10a shows the confusion matrix of the 24 tones for the first three fret and "open string" of the six strings on guitar. The 90% percentile accuracy is 94%.

We further test the sensing accuracy for blender and fan as these are commonly seen vibration/rotation sources in our lives. We bought four blenders of different brands namely KENWOOD, ACA, Bear and BARUN from the market and evaluate the frequency monitoring accuracy. As depicted in Fig. 10b, the frequency error is well within $2Hz$. We observe that the error increases with higher frequency. The reason is that the blenders experience larger displacement when the vibration frequency is higher. In addition, for fan rotation sensing, the error is kept around $1Hz$ when the tag-fan distance is 1 meter.

F. Application Studies in the Wild

To evaluate the performance of our system further, we apply RF-Ear to a real power plant for sensing. Four typical industrial equipments including Induced-draft-fan(IDF), Blowers, Vacuum-pump (V-pump) and Condensate-pump (C-pump) are selected to measure their rotating speeds with our system deployed as shown in Fig. 6d. They have a main rotating shaft (5-10 meters of length) but only a small part (about 10cm of length) is exposed. We place the tag 20 centimeters away from the exposed shaft and the antenna is deployed 3 meters away from the tag. We monitor their rotating speeds in real-time for 6 hours and compare RF-Ear's results with the ground-truths measured by the laser device.

Fig. 10 depicts the mean errors of the four equipments which are 0.56Hz, 0.68Hz, 0.45Hz and 0.50Hz and the ground truth are 50Hz, 30Hz, 45Hz, 20Hz respectively. We note that there is a significant amount of high-decibel sound noise induced by the machines during the rotation process. However, the sound noise has little impact on the performance which shows the robustness of RF-Ear. Last but not the least, the end-to-end latency of our system is only about 1 second which meets the requirements of real-time monitoring.

VII. Related Work

Sensor-based methods: Conventional contactless RPM measurement in industry is achieved with the electromagnetic RPM sensors [12]. However, each sensor can only monitor a specific group of rotors. Laser-based methods [8], [9] only work in line-of-sight scenarios and can monitor one device a time. Yang *et al*. propose VibID [33], which employs one low-cost vibration motor embedded in a wearable to vibrate and then utilizes the unique response captured by the accelerometer to identify each individual.

Camera-based sensing: Seitz *et al*. [24] propose a general framework of employing image analysis for monitoring 3-dimensional repetitive motions. Veeraraghavan *et al*. [25] can capture the high-speed rotation information with a low-frame-rate camera. They recover the signals with compressive sensing method on video frames. As for periodic motion, Laptev *et al*. [19] sense the periodic motion by exploiting the similarity of consecutive images.

RFID-based sensing: Many exciting new applications have been enabled by RFID sensing including localization and tracking [26], [27], [32], gesture and activity recognition [23], [28], breathing detection [15], liquid leakage detection [14], material and shape sensing [17], [29], vibration sensing [7], [11], [21], [31], [34]. Recently, Tagbeat [31] proposes an approach which utilizes COTS hardware to sense mechanical vibration and estimate its vibration/rotation frequency. Tagtwins [11] solves this problem by attaching two tags on the machine. Zheng *et al*. design a vibration sensing system [34]. It utilizes the temporal and phase distributions of tag readings as effective features for eccentricity detection. TagSound [21] employs the harmonic backscatter signals caused by the non-linearity of RFID tag to sense the vibration frequency, achieving an mean error of $0.37Hz$ (below $100Hz$) and $4.2Hz$ (up to $2500Hz$). However, it still needs to attach tags to the vibrating object and needs an extra receiver to capture the high-frequency harmonic signals.

VIII. Conclusion

In this work we present RF-Ear, an efficient and accurate contactless multi-object vibration frequency inspection and identification system. RF-Ear can achieve a 0.2% mean error rate with $400Hz$ and identify 4 vibrating devices with 90% accuracy. With comprehensive experiments conducted in both lab and a real-world power plant, outstanding performance is achieved. RF-Ear can be used in device trouble-shooting due to the successful detection of $2mm$ amplitude error.

Acknowledgment

Xiang-Yang Li and Panlong Yang are the co-corresponding authors. The research is partially supported by National Key R&D Program of China 2018YFB0803400, China National Funds for Distinguished Young Scientists with No.61625205, China National Natural Science Foundation with No. 61632010, 61772546, ł61751211, 61772488, ł61832010, No. 61520106007, Key Research Program of Frontier Sciences, CAS. No. QYZDY-SSW-JSC002.

REFERENCES

[1] Accelerometer. https://en.wikipedia.org/wiki/Accelerator.

[2] Alien tags. http://www.alientechnology.com/tags/2x2/.

[3] Epcglobal. Low level reader protocol.

[4] Impinj, inc. http://www.impinj.com/.

[5] Laser tachometers. https://xlraceparts.com/5-best-laser-tachometers/.

[6] Llrp toolkit. http://llrp.org/.

[7] Vibration sensor tag. https://gaorfid.com/.

[8] P. Castellini, G. M. Revel, and E. P. Tomasini. Laser doppler vibrometry. In *An Introduction to Optoelectronic Sensors*, pages 216–229. World Scientific, 2009.

[9] P. Cheng, M. S. M. Mustafa, and B. Oelmann. Contactless rotor rpm measurement using laser mouse sensors. *IEEE Transactions on Instrumentation and Measurement*, 61(3):740–748, 2012.

[10] H. Du, P. Li, H. Zhou, W. Gong, G. Luo, and P. Yang. Wordrecorder: Accurate acoustic-based handwriting recognition using deep learning. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1448–1456. IEEE, 2018.

[11] C. Duan, L. Yang, H. Jia, Q. Lin, Y. Liu, and L. Xie. Robust spinning sensing with dual-rfid-tags in noisy settings. *Energy*, 1:1–1.

[12] C. Giebeler, D. Adelerhof, A. Kuiper, J. Van Zon, D. Oelgeschläger, and G. Schulz. Robust gmr sensors for angle detection and rotation speed sensing. *Sensors and Actuators A: Physical*, 91(1-2):16–20, 2001.

[13] GuitarTuna. Url: http://www.androidtapp.com/uploads/2014/12.

[14] J. Guo, T. Wang, Y. He, M. Jin, C. Jiang, and Y. Liu. Twinleak: Rfid-based liquid leakage detection in industrial environments. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 883–891. IEEE, 2019.

[15] Y. Hou, Y. Wang, and Y. Zheng. Tagbreathe: Monitor breathing with commodity rfid systems. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 404–413. IEEE, 2017.

[16] Z. Jia, A. Bonde, S. Li, C. Xu, J. Wang, Y. Zhang, R. E. Howard, and P. Zhang. Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, SenSys '17, pages 6:1–6:14, 2017.

[17] H. Jin, J. Wang, Z. Yang, S. Kumar, and J. Hong. Wish: Towards a wireless shape-aware world using passive rfids. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 428–441. ACM, 2018.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[19] I. Laptev, S. J. Belongie, P. Perez, and J. Wills. Periodic motion detection and segmentation via approximate sequence alignment. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 816–823. IEEE, 2005.

[20] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[21] P. Li, Z. An, L. Yang, and P. Yang. Towards physical-layer vibration sensing with rfids. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 892–900. IEEE, 2019.

[22] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.

[23] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan. Rio: A pervasive rfid-based touch gesture interface. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 261–274. ACM, 2017.

[24] S. M. Seitz and C. R. Dyer. View-invariant analysis of cyclic motion. *International Journal of Computer Vision*, 25(3):231–251, 1997.

[25] A. Veeraraghavan, D. Reddy, and R. Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):671–686, 2011.

[26] C. Wang, L. Xie, K. Zhang, W. Wang, Y. Bu, and S. Lu. Spin-antenna: 3d motion tracking for tag array labeled objects via spinning antenna. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2019.

[27] J. Wang and D. Katabi. Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 51–62. ACM, 2013.

[28] J. Wang, D. Vasisht, and D. Katabi. Rf-idraw: virtual touch screen in the air using rf signals. In *ACM SIGCOMM Computer Communication Review*, volume 44, pages 235–246. ACM, 2014.

[29] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 288–300. ACM, 2017.

[30] L. Yang, Y. Li, Q. Lin, H. Jia, X.-Y. Li, and Y. Liu. Tagbeat: Sensing mechanical vibration period with cots rfid systems. *IEEE/ACM Transactions on Networking (TON)*, 25(6):3823–3835, 2017.

[31] L. Yang, Y. Li, Q. Lin, X.-Y. Li, and Y. Liu. Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 16–28. ACM, 2016.

[32] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu. See through walls with cots rfid system! In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 487–499. ACM, 2015.

[33] L. Yang, W. Wang, and Q. Zhang. Vibid: user identification through bio-vibrometry. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, page 11. IEEE Press, 2016.

[34] Y. Zheng, Y. He, M. Jin, X. Zheng, and Y. Liu. Red: Rfid-based eccentricity detection for high-speed rotating machinery. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1565–1573. IEEE, 2018.