

# Data-Driven Imbalance and Hard Particle Detection in Rotating Machinery Using Infrared Thermal Imaging

Olivier Janssens<sup>a,\*</sup>, Mia Loccufier<sup>b</sup>, Rik Van de Walle<sup>a</sup>, Sofie Van Hoecke<sup>a</sup>

<sup>a</sup>*Ghent University - imec, IDLab, Department of Electronics and Information Systems  
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium*

<sup>b</sup>*DySC Research Group, Department of Electrical Energy, Systems and Automation - Ghent University  
Technologiepark 914, 9052 Zwijnaarde, Belgium*

---

## Abstract

Currently, temperature-based condition monitoring cannot be used to accurately identify potential faults early in a rotating machines' lifetime since temperature changes are only detectable when the fault escalates. However, currently only point measurements, i.e. thermocouples, are used. In this article, infrared thermal imaging is used which –as opposed to simple thermocouples– provides spatial temperature information. This information proves crucial for the identification of several machine conditions and faults. In this paper the conditions considered are outer-raceway damage in bearings, hard-particle contamination in lubricant and several gradations of shaft imbalance. The fault detection is done using an image processing and machine learning solution which can accurately detect the majority of the faults and conditions in our data set.

*Keywords:* Condition Monitoring, Fault diagnosis, Early fault detection, Rotating machinery, Infrared imaging, Image processing, Machine learning

---

## 1. Introduction

Rotating machinery can suffer from a wide range of sub-optimal conditions, such as misalignment, load imbalance, bearing raceway faults, bearing looseness and inadequate lubrication. These faults can escalate and propagate so that bearings can overheat, wear can be increased, spalling can occur and even other components can be damaged [1]. Hence, to reduce operational costs, it is important to detect these faults as early as possible.

A first approach to detect incipient faults is vibration analysis. Vibration-based condition-monitoring systems can detect a large part of the impending rolling element bearing (REB)

---

\*Corresponding author

*Email addresses:* [odjansse.janssens@ugent.be](mailto:odjansse.janssens@ugent.be) (Olivier Janssens), [mia.loccufier@ugent.be](mailto:mia.loccufier@ugent.be) (Mia Loccufier), [rik.vandewalle@ugent.be](mailto:rik.vandewalle@ugent.be) (Rik Van de Walle), [sofie.vanhoecke@ugent.be](mailto:sofie.vanhoecke@ugent.be) (Sofie Van Hoecke)

*URL:* <http://www.ugent.be/ea/idlab> (Olivier Janssens)

failures [2]. Generally, the time between the detection of a potential fault and functional failure for vibration analysis is considered to be long, in the range of weeks to months. Hence, vibration analysis has been the method of choice when it comes to condition monitoring (CM) of rotating machinery and rolling element bearings in general. However, vibration-based fault-detection systems are not useful for detecting lubrication related problems such as under-lubrication or over-lubrication as these do not manifest themselves as new cyclic frequencies [3].

Another type of lubrication inadequacy is hard particle contamination. These particles can consist of steel fragments, sand or any other type of residue and can damage the REB and reduce the rotating machine's operational lifetime significantly. Hence, most of the largest REB manufacturers state that lubricant contamination is the major cause of REB failure before they reach their rated life [4].

To solve this problem in offshore wind turbines for example, regular offline lubricant analysis is done. Nevertheless, this is a very intrusive and hands-on condition monitoring method as it requires a technician to manually sample the lubricant, bring it on shore and subsequently analyze it. By analyzing the lubricant it is possible to categorize the contamination level according to the ISO 4406:99 or ISO 11171 if oil lubricant is used. As this is an offline procedure it is possible for this fault to escalate and damage the REB permanently. Hence, often online particle counting is additionally done using a particle counter. By the use of this sensor it is possible to know, to a certain degree, how contaminated the lubricant is in real-time so that fault escalation can be prevented. Nevertheless, these sensors need to be installed in the lubrication system, which has to be a closed-loop oil lubrication system. Also, this is not possible using grease lubrication in a closed-loop lubrication system. Furthermore, online particle counters are very expensive to install [2] and cannot be used to detect other types of faults, i.e. raceway faults or imbalance.

It is also possible to employ temperature-based condition monitoring which uses thermocouple sensors. Thermocouples are relatively cheap compared to other condition monitoring sensors and allow for temperatures to be monitored inside machines. However, using temperature-based condition monitoring there is a much shorter time span between the detection of a potential fault in a machine and the functional failure [2] as can be seen in the P-F curve in Figure 1a. The time span is often in the range of hours to days. Hence, temperature-based detection is often only useful after initial fault detection using vibration analysis. However, in recent years infrared thermal imaging (IRT) has gained noticeable attention as it allows for visual temperature monitoring. Because of the advantage of having spatial information of the heat in components, IRT has been applied in several domains for specific tasks such as inspection of cracks, isolation, subsurface moisture, corrosion, gas flow, air flow, and welding processes [5]. Furthermore, by applying image processing and machine learning, IRT imaging can be used for autonomous condition monitoring. For example, IRT-based condition monitoring has been applied for the detection of conditions in rotating machinery such as imbalance, misalignment, coupling looseness and REB damages. These faults are often considered only to be detectable using vibration-based condition monitoring. Moreover, in our previous work we showed that IRT-based automated condition monitoring can be used to detect different levels of lubrication [6]. Lubrication inadequacy such

as under-lubrication and over-lubrication is prone to cause premature REB failure, hence IRT-based condition monitoring can help to increase the time between the detection of a potential fault and the functional failure as illustrated in Figure 1b.

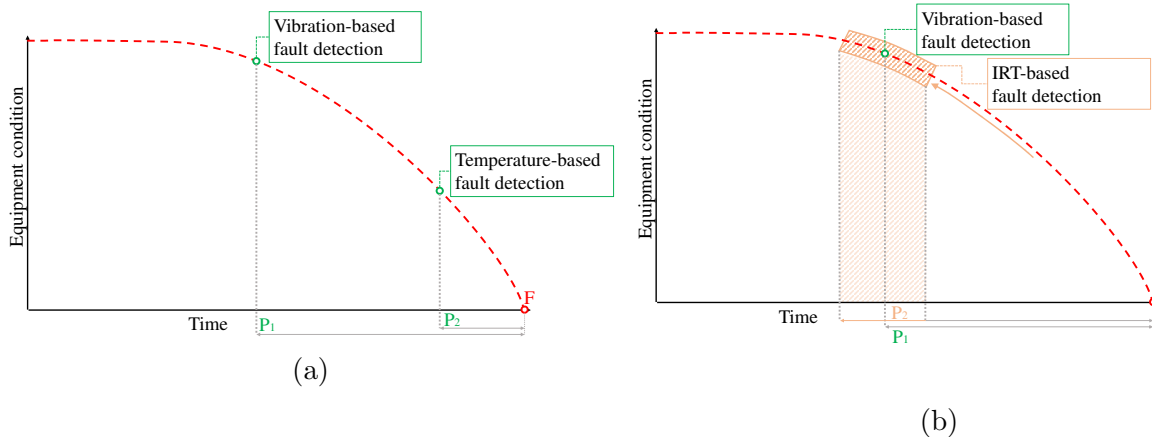


Figure 1: Potential failure and functional failure curves, illustrating (a) the time-span difference between vibration-based and temperature based fault detection, also (b) vibration-based fault detection and infrared thermal imaging based fault detection.

As lubrication contamination is either very expensive to detect, not detectable in real time or not detectable at all with current techniques, in this paper the potential of IRT imaging is investigated for automated lubricant contamination detection. Additionally, the limits of automated imbalance detection using IRT imaging are also tested –which is commonly done using vibration analysis– to illustrate that an IRT camera is a suitable additional/replacement sensor. Also, the proposed method is compared to a thermocouple-based approach as thermocouples are significantly less expensive than infrared thermal cameras and can be used to measure the temperature inside a machine. By this comparison the added value of infrared thermal imaging is further illustrated.

## 2. Related Literature

Infrared thermal imaging for the condition monitoring of rotating machinery gained noticeable attention in the last years [7]-[13]. The focus of previous research has been on the detection of conditions such as shaft misalignment, REB looseness, load imbalance and REB faults. To detect these conditions, data-driven approaches have been applied using image processing and machine learning. Often, the first step of the image processing pipeline entails the extraction of the region of interest (ROI). This is done manually or via an algorithm such as Otsu thresholding together with k-means clustering [7] or watershed-based algorithms [9]. The second step can consist of enhancing the image or ROI [10]. From this (enhanced) ROI, statistical features are derived such as the standard deviation, mean, skewness, kurtosis, variance, entropy, energy, central moments, maximum and minimum [7,

10, 11] or the components of the discrete wavelet decomposition of the thermal image [12, 13]. In the penultimate step, indiscriminating features are sometimes removed or combined to create better features. Algorithms used for this step include Principal Component Analysis [9], independent component analysis [11], discriminant analysis [10] or relief algorithm [12, 13]. The resulting features are subsequently used to determine the condition of the rotating machine using a classification algorithm. The classification algorithms which have been employed are support vector machine [11, 8], relevance vector machine [10], self-organising map [7] and linear discriminant analysis [12, 13]. Most of these approaches will result in a system that can accurately detect the latent machine condition with an accuracy of 74% up to 100 % [7]-[13].

Several remarks should be noted regarding the state-of-the-art. (1) Often the classification algorithms are trained and tested on IRT data of the same REB. Therefore, a generalized solution is not guaranteed, i.e. it cannot be confirmed that the model will work using unseen data. (2) The conditions and faults under investigation are those that can easily be detected using vibration analysis, which is a robust and widely used condition monitoring technique. However, the detection of different amounts of imbalance is still challenging, (3) especially in combination with other faults [14].

In our approach, several test-runs of different bearings are done to create disjunct training and test sets. Furthermore, several of the faults and conditions considered are difficult to identify or hardly detectable using other condition monitoring techniques. Finally, multiple conditions and faults are considered simultaneously.

### 3. Methodology

In this section we discuss the set-up which was used to create our data set, the data set itself and the proposed condition monitoring architecture.

#### 3.1. Test set-up

When designing and constructing the set-up three aspects in particular were taken into account. First, the set-up has a frame consisting of one solid piece, minimizing the amount of bolt connections. Second, the motor and bearing housings are placed in the same plane eliminating any misalignment. Third, the bolt holes are designed to minimize bolt looseness therefore guaranteeing a tight fit. More information about the set-up can be found in Table 1 and Figure 2. The information about the infrared thermal camera is summarized in Table 2. The set-up is located in a darkened room to eliminate additional noise in the IR recordings due to external influences. It should be noted that a single emissivity value is chosen for all recordings. This done because the bearing housing is made out of steel and the seal out of rubber. Both have a very similar emissivity value. Additional to the IR camera, four thermocouples are used on the set-up. Two thermocouples are put inside the REB's housing to measure the internal temperature, one is attached to the outside of the housing and the last thermocouple is placed inside the room to measure the ambient temperature.

Only the condition of the right REB is changed, which is the one further removed from the motor. In total 12 conditions are created with this set-up. An overview of these conditions

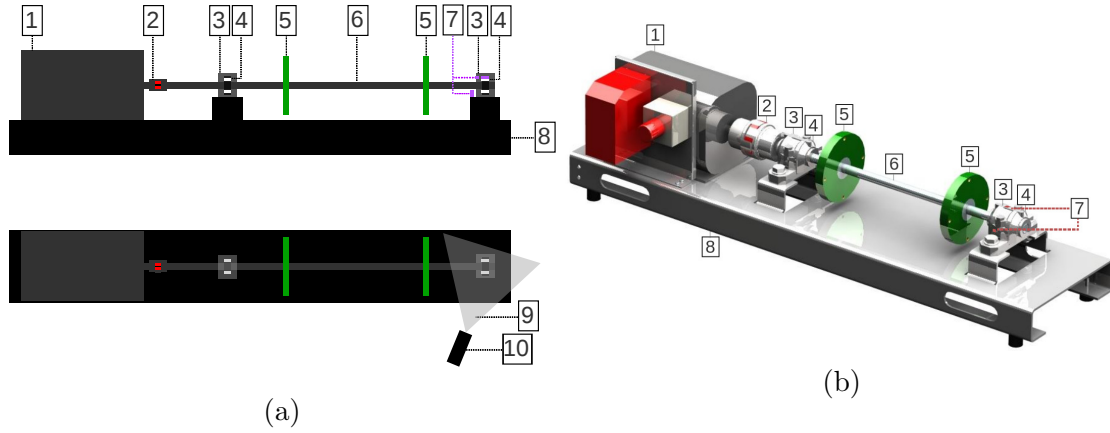


Figure 2: Side, top (a) and 3D view (b) of the set-up. The labels are: 1. servo-motor; 2. coupling; 3. bearing housing; 4. bearing; 5. disk; 6. shaft; 7. thermocouple; 8. metal plate; 9. field of view; 10. camera

Attribute	Value
Bearing code	FAG 22205-E1-K
Bearing type:	Spherical roller bearing with tapered bore and adapter sleeve
Housing code	SNV052-F-L
Housing type	Closed plummer block
Grease	Molykote BR 2 plus
Rotation speed	1500 rpm
Motor type	Single phase asynchronous induction motor
Motor power	1.1 kW

Table 1: Test set-up details

Attribute	value
Thermal camera	FLIR SC655
Capture speed	6.25 frames per second (fps)
Resolution	640 x 480 pixels
Distance: camera - housing	38 cm
Emissivity	0.9
Spectral range	7.5 - 13 $\mu\text{m}$

Table 2: Thermal camera details

Table 3: Enumeration of the 12 induced condition. Every condition consists of a bearing condition –which are healthy bearing (HB), outer-raceway fault (ORF) or hard particle contamination (HP)– but also an imbalance gradation, which are balanced, 4.1 g; 9.3 g or 13 g of added weight to the rotating rotor.

<b>Condition</b>	<b>Bearing</b>	<b>Imbalance</b>
<b>1</b>	Healthy bearing	Balanced
<b>2</b>	Healthy bearing	4.1 g or 5.5 N
<b>3</b>	Healthy bearing	9.3 g or 12.4 N
<b>4</b>	Healthy bearing	13 g or 17.3 N
<b>5</b>	Outer-raceway fault	Balanced
<b>6</b>	Outer-raceway fault	4.1 g or 5.5 N
<b>7</b>	Outer-raceway fault	9.3 g or 12.4 N
<b>8</b>	Outer-raceway fault	13 g or 17.3 N
<b>9</b>	Hard particle contamination	Balanced
<b>10</b>	Hard particle contamination	4.1 g or 5.5 N
<b>11</b>	Hard particle contamination	9.3 g or 12.4 N
<b>12</b>	Hard particle contamination	13 g or 17.3 N

can be seen in Table 3. The imbalance is created by attaching a bolt to the rotor at a radius of 5.4 cm. To introduce different levels of imbalance, a range of bolts with different weights are used. To imitate the outer-raceway fault, three shallow grooves are added mechanically to the outer-raceway of a set of REBs. Finally, to imitate the hard particle contamination 0.02 g of iron particles are mixed in the lubricant of the REBs. As lubricant, grease is used for which the required amount is calculated using Eq. (1), where  $D$  is the outer diameter of the REB and  $B$  the inner diameter [15]. For the used REBs  $D = 52$  mm and  $B = 18$  mm.

$$m = D * B * 0.0027 \text{ [g]} \quad (1)$$

All REBs contain 2.5 g of grease. Additionally, 20 grams of grease is put in the grease reservoir of the REB housing, so that the housing cavities are filled to the recommended 60% [16]. Around 170 g of grease is contained in the housing, this means that the mass fraction of the introduced particles is equal to 0.0117 %.

By means of this set-up, a data set is created incorporating the introduced conditions.

### 3.2. Data set

For every condition, five REBs are tested. Each bearing is run for one hour as part of the set-up, while the thermal camera and the thermocouples record the heating process, resulting in 5 REBs \* 12 conditions = 60 hours of recorded data. It was observed that the REB’s temperatures usually do not further increase after a duration of 50 minutes, i.e. they reach steady state [17]. Hence, only the last 10 minutes of each hour are exported to AVI video files for further processing. To reduce the size of the video files, lossless compression is applied using the H264 standard [18]. As lossless compression is used, the image quality is not reduced. The resulting videos consist of monochrome frames for which the gray values

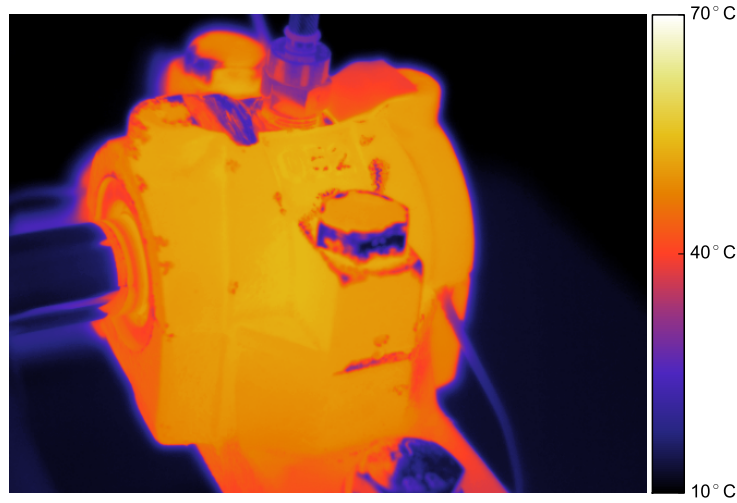


Figure 3: Pseudo-colored frame extracted from an infrared thermal video

correspond to temperatures in the range of 10 - 70° C. An example of a pseudo-colored frame can be seen in Figure 3.

### 3.3. Architecture

To automatically detect which condition is present in the machine the architecture in Figure 4 is proposed. As there are two main categories of conditions, the architecture consists of two image processing and machine learning pipelines. Pipeline one is designed to identify if the REB is healthy, contains an outer-raceway fault or has hard particles in its grease. Pipeline two is designed to determine the amount of imbalance in the system. By combining these two pipelines the 12 conditions can be identified.

### 3.4. Pipeline one

To detect if the grease of the REB contains solid particles, an outer-raceway fault or if the REB is healthy, pipeline one starts of with the preprocessing of the thermal videos. Afterwards, feature extraction and machine learning are applied.

#### 3.4.1. Preprocessing

First, sub-sampling in time is applied, resulting in a frame rate of 1 fps. The ideal frame-rate is determined experimentally for optimal results. Next, for every frame, all the pixel values are transformed to relative temperatures. This transformation is done by subtracting the ambient temperature, measured by the thermocouple, from every temperature value measured by the thermal camera. Next, the foreground, i.e. the bearing housing which is

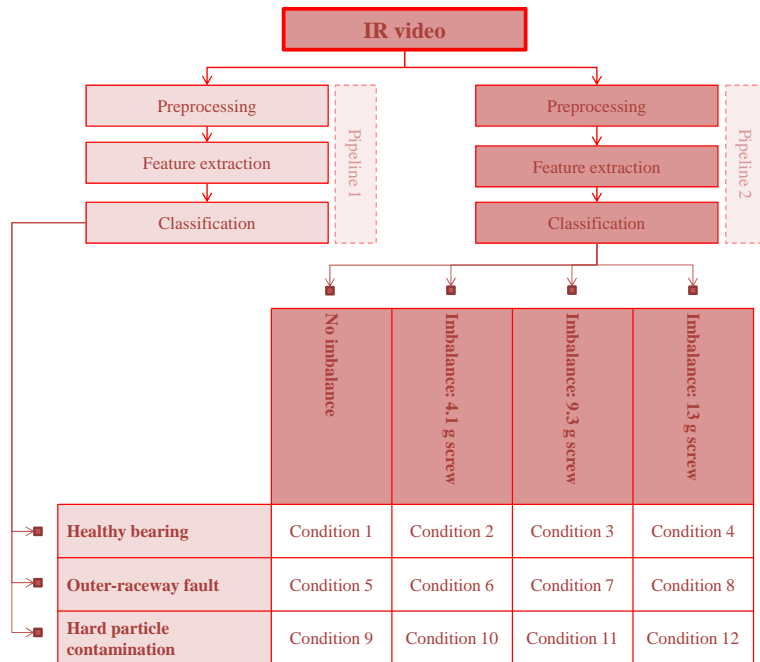


Figure 4: High-level architecture of the automatic condition monitoring system

relatively warm, needs to be isolated from the background, i.e. not-relevant colder part. This segmentation is done using a threshold determined by the Otsu algorithm [19].

In between the measurements, the IRT camera moved and so did the set-up as the measurements were done over a time-span of weeks wherein the bearings had to be swapped and the camera stored. Due to these actions, translations and rotations are induced in the data set. To enable robustness against these transformations, all the frames need to be aligned to a common reference image (i.e. image registration). By applying image registration, the REB housing will be in the same position in every frame of every recording. Image registration is applied by first extracting a random reference frame from a random recording. From this reference frame ORB key points are extracted resulting in binary descriptors [20]. Next, for every recording ORB key points are also extracted from the first frame of the recording. The goal is to find matching keypoints in both frames so that the transformation between the two sets of key points can be determined. To find matching key points, the hamming distance is used which allows for very fast matching computation compared to for example the euclidean distance [21]. Only key points for which the hamming distance is very small are kept and subsequently used to determine an affine transformation. The affine transformation describes how an image should be translated, rotated and scaled so that it will match the geometry of the reference frame. This transformation is calculated for every recording and applied to every frame, resulting in recordings wherein the housing is approximately in the same place. It should be noted that the translation and rotation



of the affine transform will not influence the pixel intensities and hence the information in the images. However, scaling can influence the information in the images. For example, down-scaling will remove a part of the information. Nevertheless, the amount of scaling required in our approach is negligible.

As last step, a boxcar filter is applied on every frame to eliminate high frequency components in the image. The goal of this step is to reduce the influence of edges in the feature extraction step as these are not informative for fault detection. Next, from these preprocessed videos, features are extracted.

### 3.4.2. Features extraction

To minimize the influence of noise in the features –inherent to the thermal camera and the environment– windowing is applied. A window contains one minute of IRT video and overlaps with its neighbouring windows by 30 seconds, resulting in 19 windows per video sequence, each window containing 30 frames. The window length can be moderately modified without influencing the end result. The goal of this step is to extract features from every frame inside a window and subsequently average the features within each respective window. In total two features are extracted from the preprocessed frames. The literature overview indicates several usable and reproducible image-processing features such as the central moments, HU moments, standard deviation, mean, skewness, kurtosis, entropy, variance,  $M_{20}$ , maximum and minimum from the image histogram. For the specific faults considered in this article, the standard deviation and the Gini coefficient proved useful.

The standard deviation (of the pixel values) describes how broad the range of the distribution of the value pixels is, and how spread-out the pixel values are in this range. The second feature is the Gini coefficient. The Gini coefficient measures inequality and is often used in economics and astronomy [22]. We recently also showed that the Gini coefficient helps to indicate the level of lubricant in a REB [6]. Both the Gini coefficient and the standard deviation are measurements of dispersion, however, as opposed to the standard deviation, the Gini coefficient is not a metric based on central tendency, i.e. deviation from the mean. To calculate the Gini coefficient, two cumulative distributions are required. The first is the cumulative distribution of the sorted pixel values of the IR image (IRCS). The second is the line of equality (LOE), which is a cumulative distribution of the image pixels values as if they were equal. Examples of these curves can be seen in Figure 5a and 5b. In these figures, the light shaded gray area represents the area under the LOE (AU-LOE) and the dark shaded gray area represents the area under the IRCS (AU-IRCS). If the pixel values of the IR image are completely uniformly distributed, i.e. if all pixel values are equal, the AU-LOE is equal to the AU-IRCS. The Gini coefficient is directly related to the size of the area in between the LOE and IRCS and therefore calculated according to equation 2. If the Gini coefficient is equal to 0 %, there is total equality, and if the Gini coefficient is equal to 100 % there is total inequality. To make sure that only the pixels of the housing are taken into account when calculating the Gini coefficient, only non-zero pixel values are used<sup>1</sup>.

---

<sup>1</sup>When comparing this work with our previous work, it should be noted that this is not done in our

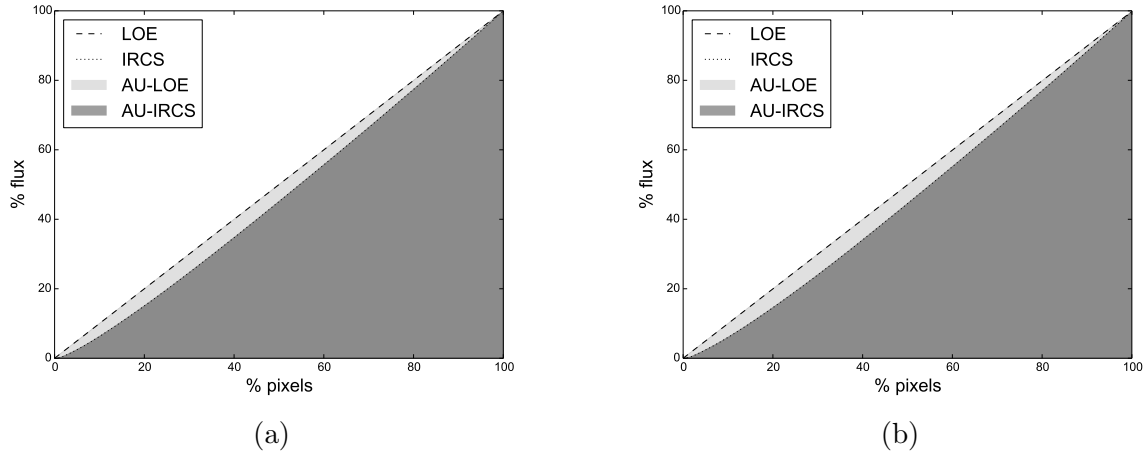


Figure 5: Graphical representation of the Gini Coefficient. The Gini coefficient for the MILB case is 0.72% higher compared to the HB case. Flux, i.e. heat flow, corresponds to the brightness of the pixels.

In the end, the design matrix provided to the classification algorithm in the next step contains 1140 samples (60 records times 19 windows) each containing two features.

$$\text{Gini coefficient} = \frac{\text{AU-LOE} \cap \text{AU-IRCS}}{\text{AU-LOE}} \quad (2)$$

### 3.4.3. Classification

To automatically identify the conditions present in industrial rotating machinery using the extracted features, a classifier is required. From the data and previous work [6], it can be seen that the classification problem is a non-linear classification problem, i.e. the classes are not possible to divide using a straight line. Furthermore, there are multiple classes to distinguish making the problem a multi-class classification problem. Several classification algorithms fulfill these requirements such as k-nearest neighbors, support vector machines and random forest classifiers. Tests were done using these algorithm and as in [6], it was determined that the random forest classifier (RFC) provides the best results. The RFC in this pipeline contains 100 decision trees, for which the amount was empirically determined using grid search.

### 3.5. Pipeline two

As is depicted in Figure 4, the goal of pipeline two is to determine if there is imbalance in the system and in if there is imbalance, determine how much. To do this, as in pipeline one, first the videos are preprocessed.

---

previous work [6]. However, by not taking the zero pixels into account the Gini coefficient is merely scaled compared to when the zero pixels are taken into account.

### 3.5.1. Preprocessing

First, as consecutive frames contain redundant information, sub-sampling in time is applied. Evaluation showed that a frame rate of 0.5 fps provides optimal results. Subsequently, to minimize the influence of noise and outliers in the feature extraction step, windowing is applied here too.

### 3.5.2. Feature extraction

As described in our previous work [6], due to imbalance the set-up will vibrate. Although, this vibration is hardly noticeable, it can be seen in the video when differencing two consecutive frames. For example when subtracting Figure 6a from Figure 6b, Figure 6c is obtained showing the outline of the housing as a result of the movement due to imbalance. A threshold is additionally applied to remove background noise in the differenced image as is displayed in Figure (6d). The optimal threshold value is empirically determined. Note that only the right side of the image is used as the movement of the shaft, i.e. rotation, is not related to the imbalance condition. Hence, if the shaft is included in the image lower classification performance is achieved.

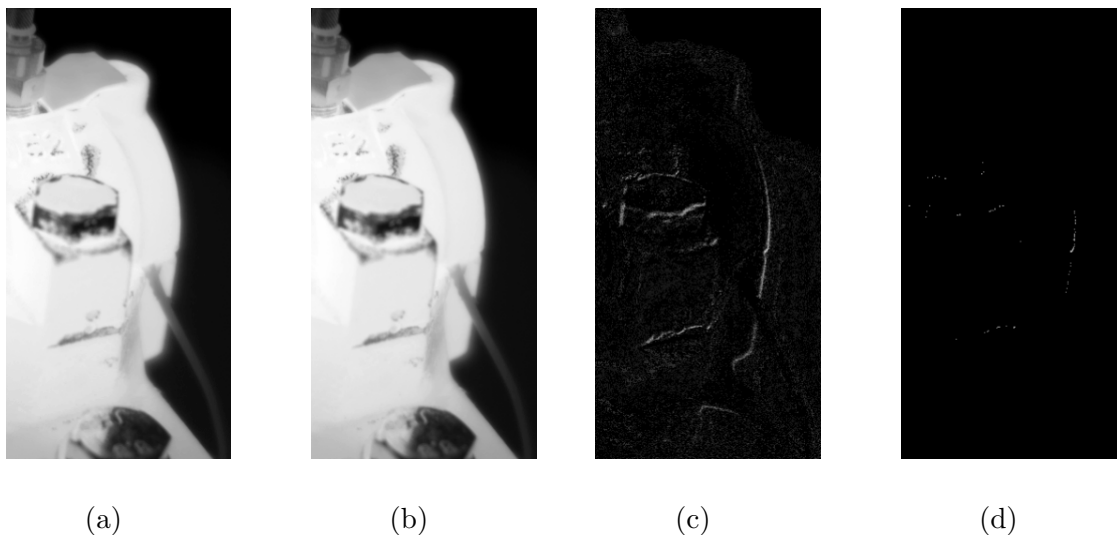


Figure 6: Movement detection. Parts of consecutive frames (a)(b) are differentiated resulting in (c). Afterwards a threshold is applied (d).

To detect the degree of imbalance, the difference between frames need to be quantified as features for the classification. Therefore, the histogram of the pixels is constructed along the x-axis and y-axis. Examples of these histograms can be seen in Figure 7a and 7b. Not only are the number of pixels important, but also the width of the histograms. Generally, if there is imbalance, the histograms will be wider. To describe this distribution, the standard deviation (SD) is chosen as it captures how spread out a distribution is. Also, the kurtosis is used as it describes how peaked a distribution is.

When the SDs of the histograms along both axis are plotted four regions corresponding to the four balance/imbalance conditions can be observed (Figure 8a). It can be stated

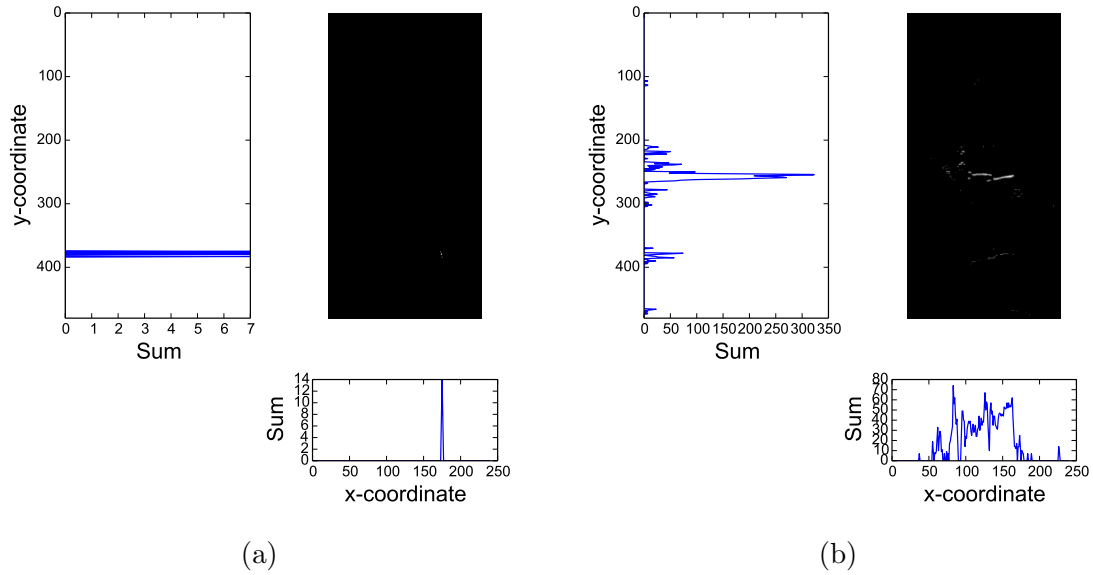


Figure 7: Examples of the histogram along x and y axis for a healthy bearing when in balance (a) and imbalance (b).

that the higher the SDs, the more imbalance is present in the system. Nevertheless, the clusters of points do partially overlap. Generally, a classifier will already be able to make a good distinction between the different levels of imbalance using the SDs, however, by adding the kurtosis the classifier's performance improves. In the end, the design matrix provided to the classification algorithm in the next step contains 1140 samples (60 records times 19 windows) each containing four features.

### 3.5.3. Classification

As in pipeline one, a random forest classifier (RFC) is also used for pipeline two. When plotting the decision boundaries that the RFC determined, the regions which correspond to the level of imbalance, become visible (Figure 8b). Tests were also conducted using support vector machines using linear, polynomial and radial basis kernels. Nevertheless, these performed less well compared to the RFCs.

## 3.6. Thermocouple-based fault detection

It can be questioned if the added value of using an (expensive) thermal camera compared to using (cheap) thermocouple sensors is justifiable. Hence, to put the results of the thermal camera based system in perspective, a system which only uses thermocouple measurements is discussed here.

### 3.6.1. Data gathering and preprocessing

Each one of the thermocouples records a temperature value every second. As with the IRT data, only the last 10 minutes of the data is used, which is recorded during the steady-state period. It is observed that only the thermocouples which were placed inside the housing

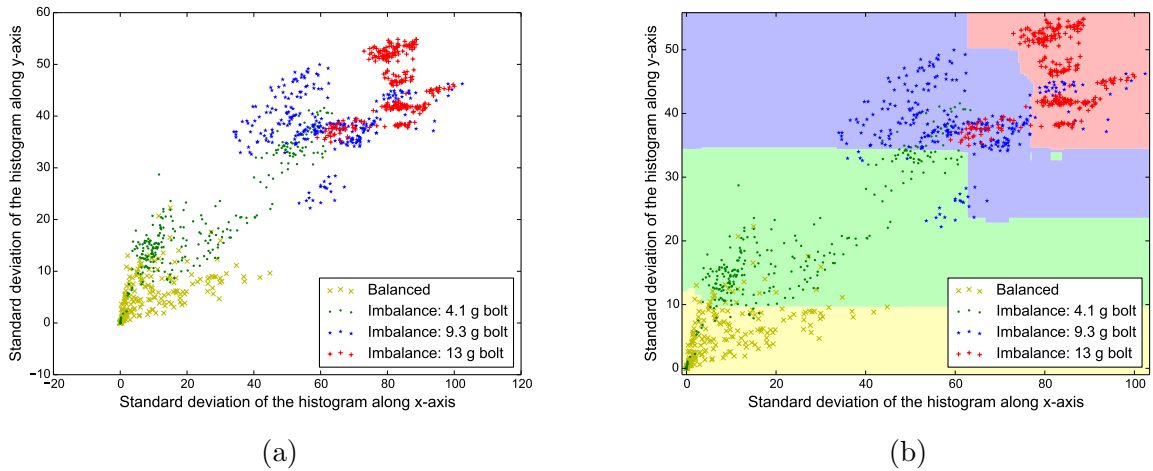


Figure 8: Scatter plots of the standard deviations along the image axis for the different levels of imbalance  
a. When these samples are classified, clear, logical regions are noticeable according to the level of imbalance  
b.

provide useful data. As preprocessing step, the signals from these two thermocouples are corrected for the ambient temperature as in Eq. 3.

$$\mathbf{t}_{relative} = \mathbf{t}_{absolute} - \mathbf{t}_{ambient} \quad (3)$$

### 3.6.2. Feature extraction and classification

From the two corrected signals eight features in total are extracted, i.e. the standard deviation, the mean, the median and  $\Delta t$  which denotes the the difference between the highest recorded temperature and the lowest recorded temperature. These four features are extracted for the two signals resulting in 8 features. After the feature extraction the samples are used by a random forest classifier to automatically distinguish between the different conditions.

## 4. Results

In this section, the results of the IRT-based system are discussed. To put the results in perspective, the thermocouple-based results are also presented.

### 4.1. Evaluation metrics

To quantify the performance of the different systems, four error measurements are used: accuracy, precision, recall and f1-score for which the formulas can be seen in Eqs. 4–7, with  $|TP|$  being the amount of true positive classifications;  $|TN|$ , the amount of true negative classifications;  $|FP|$ , the amount of false positive classifications, e.g. a false alarm, and  $|FN|$ , the amount of false negative classifications, e.g. missed faults. These different metrics are chosen because they reflect the impact on CM decisions. Suppose a CM system triggers an

alarm to alert an operator when the classifier supposedly detects a defect. Then, it is more interesting to be alerted of all the faults, even if among the fault there are some false alarms. Nevertheless, the operator doesn't want to have too many false alarms since this increases the operational cost due to unnecessary downtime. In other words, if many alarms are triggered, a lot of faults are brought to the operator's attention (higher recall), nevertheless, there are also more false alarms (lower precision). On the other hand, if only real faults are flagged, and some are missed and there are no false alarms, there will be a high precision, but a low recall. A good classifier will maximize both precision and recall, so that:

1. An alarm is only triggered when there is an actual fault.
2. No faults are missed.
3. There are no false alarms.

These requirements are indirectly enforced when optimizing the f1-score. Also, the accuracy is chosen since its ease of interpretation as it is the ratio between the amount of correctly classified samples and the total amount of samples.

$$accuracy = \frac{|TP| + |TN|}{|TP| + |FP| + |FN| + |TN|} \quad (4)$$

$$precision = \frac{|TP|}{|TP| + |FP|} \quad (5)$$

$$recall = \frac{|TP|}{|TP| + |FN|} \quad (6)$$

$$f1 - score = 2 * \frac{precision * recall}{precision + recall} \quad (7)$$

#### 4.2. Evaluation procedure

To objectively evaluate the performance of the systems, five-fold cross-validation is used. Hence, from the 60 recordings, 48 recordings, are used to train the system, and 12 recordings, i.e. one recording for each condition taken from the same REB, are used to test the system. This procedure is done five times so that every REB is used only once for testing. As one REB is left out the training set, this method can be seen as leave-one-bearing-out cross-validation. It is important to note here that the system does not use data of a hold-out REBs during training in order to avoid overfitting and have a generalized model.

#### 4.3. Pipeline one

As listed in Table 4, the IRT based fault detection system achieves better results (on average 5 % better) than the thermocouple-based system. From the confusion matrix of the IRT based system (Figure 9a) and the confusion matrix of the thermocouple based system (Figure 9b) it can be seen that the different classes can be detected, to a certain degree, by both the IRT-based system and the thermocouple-based system. The fact that the thermocouple-based results are in line with the IRT-based results confirm that temperature

is a crucial element in detecting these machine faults. However, although the results of the IRT-based system are slightly better than the results of the thermocouple-based system, the added value of using a thermal camera is still low when taking the cost price of a thermal camera into account as similar results can be achieved using cheaper thermocouple sensors.

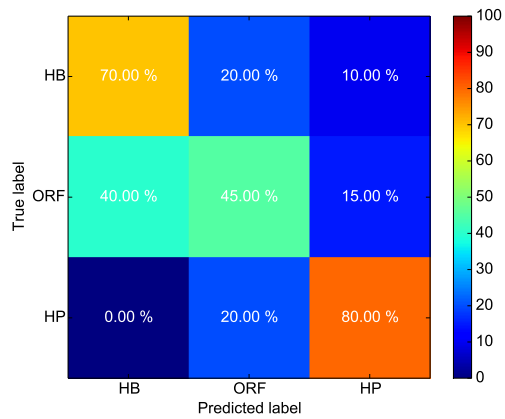
From the confusion matrices (Fig. 9a and Fig. 9b) of the classifiers, it can be seen that, in general, the major problem is the detection of the outer-raceway faults. This problem is also visible in the scatter plots of the features extracted from the infrared thermal video (Figure 10). The scatter plots show that there is a lot of overlap between the HB samples and ORF samples and also a lot of overlap between the HP samples and the ORF samples. When the samples of the ORF are removed (Figure 10b), two distinct clusters of data can be seen. Also, when the ORF samples are removed from the data set, the classifiers perform much better as summarized in Table 5. The IRT-based system’s accuracy rises by 27.5 % and the thermocouple based system’s accuracy by 15 %. Also, the standard deviations decline, indicating the reduced variability in the classifiers’ performance. The confusion matrices (Figure 9c and 9d) also show that the remaining classes, i.e. HB and HP, are better distinguishable, especially when using an infrared thermal camera. From these experiments it can be concluded that by using temperature data, outer-raceway faults are almost not detectable, neither by using IRT data nor thermocouple data. Contrary, the detection of hard particles is very feasible, especially when using infrared thermal data.

Table 4: Performance results of the IRT-based system and the thermocouple-based system for the detection of healthy REBs, REBs with an ORF and REBs with HP contamination

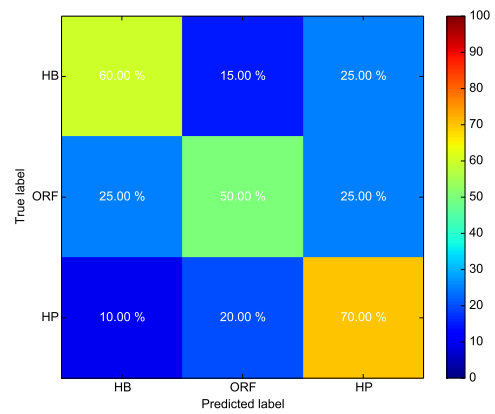
<b>Metric</b>	<b>IRT-based pipeline one</b>	<b>Thermocouple-based pipeline one</b>
<b>Accuracy</b>	65 % ( $\sigma = 16.16$ %)	60 % ( $\sigma = 21.34$ %)
<b>Precision</b>	66.69 % ( $\sigma = 16.61$ %)	63.78 % ( $\sigma = 24.72$ %)
<b>Recall</b>	65 % ( $\sigma = 16.16$ ,%)	60 % ( $\sigma = 21.34$ %)
<b>F1-score</b>	65.84 % ( $\sigma = 16.38$ %)	61.83 % ( $\sigma = 22.91$ %)

Table 5: Performance results of the IRT-based system and the thermocouple-based system for the detection of healthy REBs and REBs with HP contamination

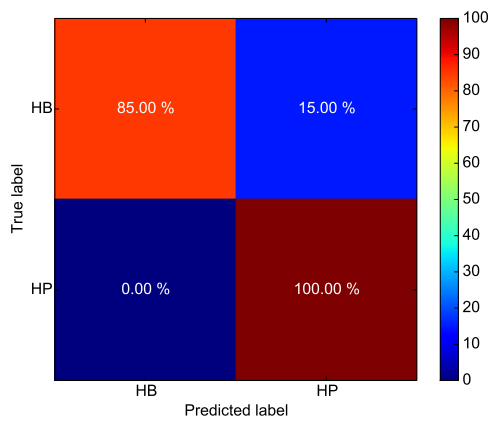
<b>Metric</b>	<b>IRT-based pipeline one</b>	<b>Thermocouple-based pipeline one</b>
<b>Accuracy</b>	92.5 % ( $\sigma = 10$ %)	75 % ( $\sigma = 17.68$ %)
<b>Precision</b>	94.67 % ( $\sigma = 6.86$ %)	78.71 % ( $\sigma = 16.85$ %)
<b>Recall</b>	92.5 % ( $\sigma = 10$ %)	75 % ( $\sigma = 17.68$ %)
<b>F1-score</b>	93.57 % ( $\sigma = 8.14$ %)	76.81 % ( $\sigma = 17.26$ %)



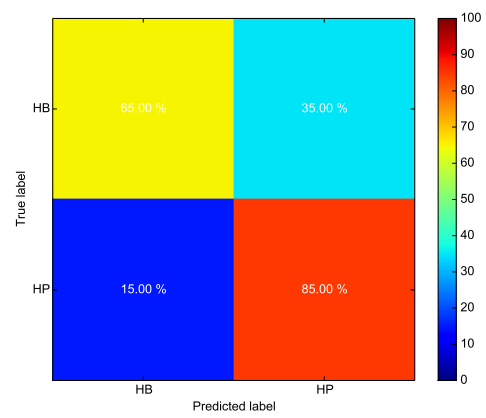
(a)



(b)



(c)



(d)

Figure 9: The confusion matrices of the classifier using IRT data (a)(c) and thermocouple data (b)(d) for the classification between a healthy bearing, (a bearing with an outer-raceway fault) and a bearing with hard-particle contamination.



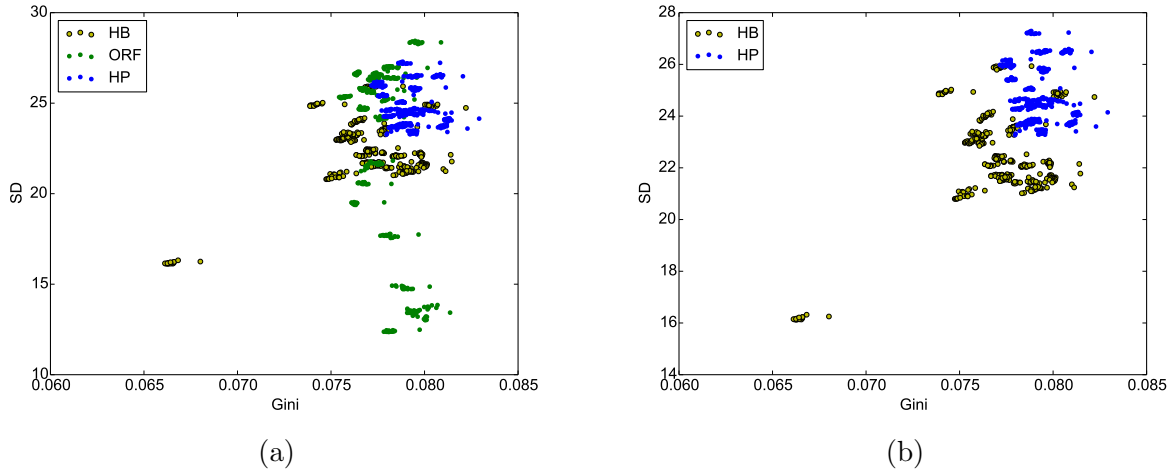


Figure 10: Scatter plots of the samples of the healthy bearings, the bearings with an outer-raceway fault and bearings with hard particle contamination (a) and a scatter plot of the samples of the healthy bearings and the bearings with hard particle contamination (b)

#### 4.4. Pipeline two

The detection of the amount of imbalance is not feasible using thermocouples, which is to be expected as thermocouples do not capture movement effects. However, the thermal camera does capture this movement effect and achieves good results in the detection of different gradations of imbalance as can be observed from Table 6. When visualizing the confusion matrix (Figure 11) of the classification results obtained using IRT data it can be seen that classifier can confuse a certain imbalance condition with another imbalance condition for which the weight difference is small. It can be concluded that there is a relationship between the standard deviation feature extracted from the IRT data and the amount of imbalance. In Appendix A the relationship between the amount of imbalance and the observable displacement in the image is described analytically and quantified. This quantification also illustrates the limits of optical imbalance detection.

Table 6: Performance results of the IRT-based system and the thermocouple-based system for the detection different levels of imbalance

Metric	IRT-based pipeline two	Thermocouple-based pipeline two
<b>Accuracy</b>	88.33 % ( $\sigma = 12.47$ %)	23.33 % ( $\sigma = 14.34$ %)
<b>Precision</b>	90.42 % ( $\sigma = 10.26$ %)	21.42 % ( $\sigma = 14.3$ %)
<b>Recall</b>	88.33 % ( $\sigma = 12.47$ %)	23.33 % ( $\sigma = 14.34$ %)
<b>F1-score</b>	89.36 % ( $\sigma = 11.26$ %)	22.33 % ( $\sigma = 14.34$ %)

#### 4.5. Combination

As the goal is to have a system that can distinguish between the 12 conditions in Table 3, the results of two pipelines using IRT data are combined to get single fault/condition

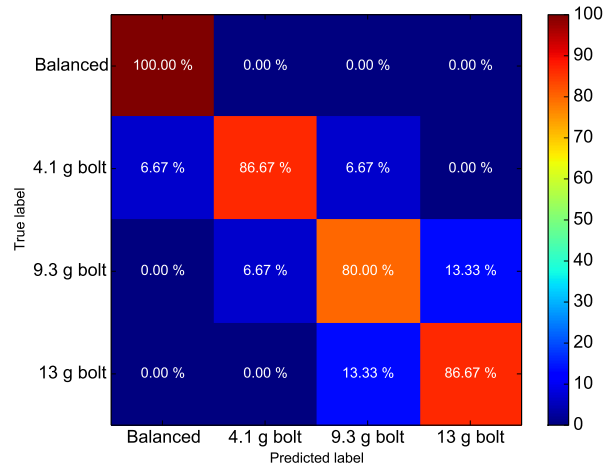


Figure 11: Confusion matrix of the imbalance detection.

predictions. The results of the combined pipelines can be seen in Table 7. As can be expected, the results when outer-raceway faults are present in the data set are not good. Nevertheless, when the outer-raceway faults are removed, the remaining 8 conditions/faults can be classified significantly more accurately.

Two aspects should be noted here. First, it is only feasible to detect multiple faults at the same time using the IRT imaging approach as it is not possible to detect gradations of imbalance using thermocouples. Second, IRT works very well for the detection of lubrication related faults and imbalance, but for other faults such as outer-raceway-faults, vibration analysis is more suitable and will be required for a robust multi-sensor system that can deal with a broad range of faults.

Table 7: Performance results when combining the two pipelines that use infrared thermal data.

Metric	All condition/faults	All conditions except for ORFs
<b>Accuracy</b>	55.0 % ( $\sigma = 10.00$ %)	82.5 % ( $\sigma = 12.47$ %)
<b>Precision</b>	41.94 % ( $\sigma = 14.25$ %)	84.76 % ( $\sigma = 12.47$ %)
<b>Recall</b>	55.0 % ( $\sigma = 10.00$ %)	82.5 % ( $\sigma = 12.47$ %)
<b>F1-score</b>	47.54 % ( $\sigma = 12.28$ %)	83.61 % ( $\sigma = 12.47$ %)

## 5. Conclusions and future work

In this article infrared thermal imaging is used in combination with image processing and machine learning for the detection of 12 conditions and faults in rotating machinery. The set of conditions and faults contain several levels of shaft imbalance, outer-raceway faults in bearings, hard-particle contamination of the lubricant and healthy bearings. A set of different bearings were used to create a large data set on which the image processing

and machine learning techniques were applied. When distinguishing between the 12 different conditions/faults poor performance is achieved (55 % accuracy). It is determined that this is due to the fact that outer-raceway faults are very hard to detect using the proposed method based on infrared thermal imaging data. When the outer-raceway faults are removed from the data set the accuracy rises to 82.5 %. Coincidentally, raceway faults are generally easily detectable using vibration analysis.

To put the results achieved using infrared thermal imaging in perspective, a similar analysis is performed using thermocouple data. The results indicate that thermal imaging enables better results as identifying different amounts of imbalance requires spatial information, which is not possible using thermocouples.

It can be concluded that infrared thermal imaging is a promising technique for automated condition monitoring as it enables the detection of difficult and complex conditions non-intrusively. Furthermore, infrared thermal imaging can extend the time between the detection of a potential fault and the functional failure as conditions can automatically be detected which are usually not detectable in an online-manner or only detectable using vibration analysis. In future work infrared thermal imaging will be combined with other sensors such as accelerometers in a multi-sensor system with the goal of accurately detecting a multitude of faults to further decrease the time between the potential fault and the functional failure.

## 6. Acknowledgements

This work was funded by the O&M Excellence project, a VIS project of the Institute for the Promotion of Innovation through Science and Technology in Flanders (VLAIO), and has been performed in the framework of the Offshore Wind Infrastructure Application Lab (<http://www.owi-lab.be>).

## Appendix A. Imbalance displacement

As a result of imbalance, the set-up will vibrate. This vibration has a certain displacement that needs to be measurable by the camera. Hence, in this appendix this displacement is calculated for the three imbalance conditions. Next, the area size that is captured in each pixel is calculated. Finally, the displacement is converted into pixels.

The following parameters are known of the set-up (Figure 2).

- Mass of the shaft: 1.618 kg
- Mass of one disks: 2.317 kg
- Total mass ( $m$ ) = 6.252 kg
- Rotation frequency ( $\omega$ ): 25 Hz

- First eigenfrequency of the set-up ( $\omega_n$ ): 17.8 hz. This was determined using an impact test.
- Radius at which the screws were attached to the disk ( $r$ ) =  $5.4 \times 10^{-2}$  m
- Mass of screw 1:  $4.1 \times 10^{-3}$  kg
- Mass of screw 2:  $9.3 \times 10^{-3}$  kg
- Mass of screw 3:  $13.0 \times 10^{-3}$  kg

The displacements are calculated according to equation A.1. The right-hand side of the denominator can be ignored due to the fact that there is no, or no significant, dampening, hence  $\zeta$  will be zero or very close to zero.

$$q_0 = \frac{\frac{m_r r \omega_n^2}{m \omega_n^2}}{\sqrt{(1 - \frac{\omega^2}{\omega_n^2})^2 + 4\zeta^2 \frac{\omega^2}{\omega_n^2}}} \quad (\text{A.1})$$

The resulting displacements are:

- Displacement by screw 1: 73.15  $\mu\text{m}$
- Displacement by screw 2: 165.9  $\mu\text{m}$
- Displacement by screw 2: 231.9  $\mu\text{m}$

Next, the area size that is captured by a single pixel, given the properties of the camera, is determined.

Two properties are required:

- Vertical field-of-view ( $v$ ):  $19^\circ$
- Distance from the camera to the bearing housing ( $d$ ): 380 mm

Using equation A.2, it is determined that at a distance of 380 mm, the camera can capture 130.84 mm vertically. By dividing this vertical dimension by the amount of vertical pixels ( $\frac{130.84 \text{ mm}}{480 \text{ pixels}}$ ) it is determined that 1 pixel = 272.59  $\mu\text{m}$ .

$$\text{Vertical dimension} = d \tan(v) \quad (\text{A.2})$$

Finally, the displacements are converted into pixels (equations A.3, A.4, A.5). Note that displacements are multiplied by two as displacements are both in the upwards and downwards direction.

$$s_1 = 2 \frac{73.15 \mu\text{m}}{272.59 \frac{\mu\text{m}}{\text{pixel}}} = 0.54 \text{ pixels} \quad (\text{A.3})$$

$$s_2 = 2 \frac{165.9 \text{ } \mu\text{m}}{272.59 \frac{\mu\text{m}}{\text{pixel}}} = 1.22 \text{ pixels} \quad (\text{A.4})$$

$$s_3 = 2 \frac{231.9 \text{ } \mu\text{m}}{272.59 \frac{\mu\text{m}}{\text{pixel}}} = 1.70 \text{ pixels} \quad (\text{A.5})$$

In general, there is a relationship between the weight causing the imbalance and the displacement in the image<sup>2</sup>. This also means that there is a limit to the amount of imbalance that can be distinguished using this technique, as amounts of imbalance that are almost the same will not be distinguishable from one another. Furthermore, due to noise, inherent to a thermal camera, very low levels of imbalance will not be measurable with great certainty.

It should also be noted that due to this relationship it should also be possible to calculate the weight of imbalance to a certain degree, given the measured displacement. However, this requires further investigation.

## References

- [1] SKF, Bearing self study guide, <http://www.skf.com/binary/12-69177/457640.pdf>, 2008. Accessed: 2016-02-15.
- [2] K. Fischer, F. Besnard, L. Bertling, Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience, *Energy Conversion, IEEE Transactions on* 27 (2012) 184 – 195.
- [3] P. Bokoski, J. Petrovi, B. Musizza, ani Jurii, Detection of lubrication starved bearings in electrical motors by means of vibration analysis, *Tribology International* 43 (2010) 1683 – 1692.
- [4] D. Koulocheris, A. Stathis, T. Costopoulos, G. Gyparakis, Comparative study of the impact of corundum particle contaminants size on wear and fatigue life of grease lubricated ball bearings, *Modern Mechanical Engineering* (2013).
- [5] S. Bagavathiappan, B. Lahiri, T. Saravanan, J. Philip, T. Jayakumar, Infrared thermography for condition monitoring – a review, *Infrared Physics & Technology* 60 (2013) 35 – 55.
- [6] O. Janssens, R. Schulz, V. Slavkovikj, K. Stockman, M. Loccufier, R. Van de Walle, S. Van Hoecke, Thermal image based fault diagnosis for rotating machinery, *Infrared Physics & Technology* 73 (2015) 78 – 87.
- [7] A. Widodo, D. Satrijo, T. Prahasto, G.-M. Lim, B.-K. Choi, Confirmation of Thermal Images and Vibration Signals for Intelligent Machine Fault Diagnostics, *International Journal of Rotating Machinery* (2012) 1–10.
- [8] G.-M. Lim, D.-M. Bae, J.-H. Kim, Fault diagnosis of rotating machine by thermography method on suport vector machine, *Journal of Mechanical Science and Technology* 28 (2014) 2948–2952.
- [9] H. Fandino-Toro, O. Cardona-Morales, J. Garcia - Alvarez, G. Castellanos-Dominguez, Bearing Fault Identification using Watershed-Based Thresholding Method, in: G. Dalpiaz, R. Rubini, G. D’Elia, M. Cocconcelli, F. Chaari, R. Zimroz, W. Bartelmus, M. Haddar (Eds.), *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*, Lecture Notes in Mechanical Engineering, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 137–147.
- [10] V. T. Tran, B.-S. Yang, F. Gu, A. Ball, Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis, *Mechanical Systems and Signal Processing* 38 (2013) 601–614.

---

<sup>2</sup>Note that several properties of the set-up and camera have to be taken into account.

- [11] G.-m. Lim, Y. Ali, B.-s. Yang, The Fault Diagnosis and Monitoring of Rotating Machines by Thermography, in: J. Mathew, L. Ma, A. Tan, M. Weijnen, J. Lee (Eds.), *Engineering Asset Management and Infrastructure Sustainability*, Springer London, 2012, pp. 557–565.
- [12] A. Younus, B. Yang, Wavelet co-efficient of thermal image analysis for machine fault diagnosis, in: *International Conference on Mechanical Engineering*, pp. 1–7.
- [13] A. M. Younus, B.-S. Yang, Intelligent fault diagnosis of rotating machinery using infrared thermal image, *Expert Systems with Applications* 39 (2012) 2082–2091.
- [14] J. Lacey, An overview of bearing vibration analysis, *Maintenance & Asset Management* 23 (2008) 32–42.
- [15] SKF, Spherical roller bearings, Online, 2009.
- [16] Schaeffler, Fag split plummer block housings of series snv, Online, 2015.
- [17] R. Schulz, S. Verstockt, J. Vermeiren, M. Loccufier, K. Stockman, S. Van Hoecke, Thermal imaging for monitoring rolling element bearings, in: *Quantitative InfraRed Thermography*, pp. 1–9.
- [18] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the h. 264/avc video coding standard, *Circuits and Systems for Video Technology, IEEE Transactions on* 13 (2003) 560–576.
- [19] O. Nobuyuki, A threshold selection method from gray-level histograms, *Systems, Man and Cybernetics, IEEE Transactions on* 9 (1979) 62–66.
- [20] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, Orb: An efficient alternative to sift or surf, in: *Proceedings of the 2011 International Conference on Computer Vision*, pp. 2564–2571.
- [21] F. Selka, V. Agnus, S. Nicolau, A. Bessaid, L. Soler, J. Marescaux, M. Diana, *Biomedical Image Registration: 6th International Workshop, WBIR 2014*, London, UK, July 7-8, 2014. Proceedings, Springer International Publishing, pp. 114–123.
- [22] J. M. Lotz, J. Primack, P. Madau, A New Nonparametric Approach to Galaxy Morphological Classification, *Astronomical Journal* 128 (2004) 163–182.