

## A Feature Selection and Extraction Method from Time-Frequency Images

Dorel Aiordăchioaie

”Dunarea de Jos” University of Galati  
Electronics and Telecom. Department  
Galati, Romania

e-mail: Dorel.Aiordachioaie@ugal.ro

Theodor D. Popescu

National Institute for Research  
and Development in Informatics  
Bucharest, Romania

e-mail: Theodor.Popescu@ici.ro

Bogdan Dumitrascu

”Dunarea de Jos” University of Galati  
Electronics and Telecom. Department  
Galati, Romania

e-mail: Bogdan.Dumitrascu@ugal.ro

**Abstract**—Time-frequency image processing is considered in the context of change detection and diagnosis purposes based on signal processing paradigm. A method for selection and extraction of features from time-frequency is considered and evaluated. New images are obtained by applying a criterion based on the contours generated by the main components of the analyzed time-frequency image. The transformed images are less complex and could be white and black only. Features based on statistical moments are considered, selected and used to define discriminant functions, in order to improve the results of the classification. The features include the number of the contours, the average area defined by the contours, the variance of the areas and the Renyi entropies. As case study, signals coming from vibration generated by faults in bearings are considered. The main output of the paper is the method of the feature selection and extraction from time-frequency images.

**Keywords** - signal; image; time-frequency transform; signal processing; feature selection; classification.

### I. INTRODUCTION

Incipient fault detection in mechanical processes by vibration monitoring is an important activity for various goals such as predictive maintenance, safety and product cost optimization. The paper presents a method for feature selection and extraction from time-frequency images of the vibration signals generated by various faults in the bearings of the rotating machines for classification purposes mainly and represents an extended and enhanced version of [1]. More facts and references are presented in the state of the art, the structure of the method is introduced, more details from the selection process of the features, and more results from experiments are presented also.

An important activity in industry, for safe work and quality of the products, is the Change Detection and Diagnosis (CDD) in various processes. These two activities are parts of a wider domain, called condition-based and predictive maintenance, as described in some excellent books with theory and applications [2] [3] [4]. In the field of vibrational processes, i.e., processes that generate mechanical vibrations, with or without faults or damages, advanced signal processing algorithms are intensively used to elaborate accurate and robust algorithms for process diagnosis [5] [6] [7].

One of the more complex signal processing method is based on time-frequency transform, and next on time-

frequency images, as described in [8] [9] [10]. The structure of such processing chain is presented in Figure 1. Signals from the process under study are pre-processed both in continuous and discrete time, mainly by filtering and scaling. Next, a time sliding window is considered for the computation of the time-frequency transform.

The parameters of the sliding window depend on the statistical properties of the analyzed signals, to meet the condition of the statistical stationarity. The coefficients of the time-frequency transform are considered as elements of an image. From this point all processing steps are based on image processing, for various tasks, as fault detection and diagnosis. Finally, from methodological point of view a set of papers and practical examples are available as [11] [12].

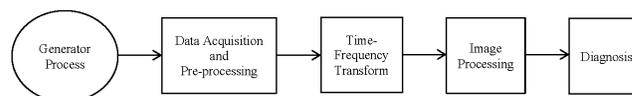


Figure 1. The block structure of signal processing for CDD

This work considers the last block before diagnosis, i.e., image processing for classification purposes. The main activities are related to the selection and extraction of the right features, in order to recognize the difference among various images. Some pre-processing steps should be considered also, as image scaling and registration.

The main processing blocks are mainly for signal processing, i.e., data acquisition and pre-processing, time-frequency transforms, image processing, classification and diagnosis, and are described in the Sections of the paper. Direct classification of time-frequency images does not offer always the best results in CDD activities, as described in [13] [14]. It is the main objective of this paper to define algorithms for feature selection and extraction, in order to obtain better results in the future classification and diagnosis stages.

The rest of this paper is organized as follows. Section II describes the basic transforms applied to the vibrating signals, i.e., time-frequency and Renyi entropy. Section III describes the basic structure of the proposed method, including data description and time-frequency images. Section IV goes into the results of the experiments, where the main results and examples are presented and discussed. The conclusion and acknowledgement close the article.

## II. DATA TRANSFORMS

The signal under transform is generated by a sliding window with a length depending on the dynamic properties of the analyzed signal, as in the Figure 2. Signal transforms are used to compute specific features of the analyzed signal or to change the analysis system, e.g., time-frequency transforms, or to compute and extract other relevant features, e.g., the Renyi entropy. The final goal of these processing steps is to detect changes in the data stream, preferably associated with individual or mixed faults of the elements and components of the process.

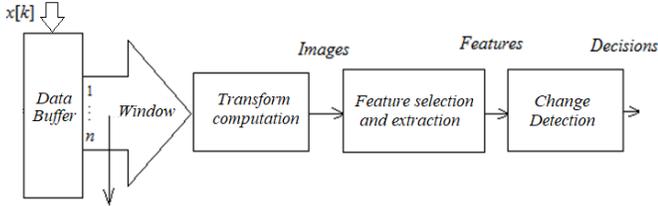


Figure 2. The structure of the processing for feature selection

### A. Time-Frequency Transform

Time-frequency transforms are advanced processing techniques for data processing, and especially for data coming from non-stationary signals. A general theoretical framework is presented in [15] [16]. Examples of signals and applications are audio signals [17], mechanical vibrations [18] or biomedical signals [19].

There are three main methods currently used for time-frequency representation and analysis. These are: (i) Short-Time Fourier Transform (STFT); (ii) Wavelet Transform (WT); (iii) Cohen class.

The STFT of a signal  $x(t) \in L^2(\mathbf{R})$  considers a window  $w(t)$ , as

$$STFT_{xw}(t, f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau} d\tau \quad (1)$$

where  $w(t)$  is the weighting window. The squared modulus is called spectrogram, as

$$S_{xw}(t, f) = |STFT_{xw}(t, f)|^2 \quad (2)$$

and constitutes a signal energy distribution in the time-frequency plane. Even the spectrogram constitutes one of the widely used methods for the analysis of non-stationary signals, in some case it is unsuitable for the compromise needed of time and frequency resolutions, i.e., it is not possible to simultaneously have good time resolution and good frequency resolution. Consequently, the user must correctly choose the characteristics of the analysis window depending on the signal structure, considering especially the proximity and evolution of the signal components in time and in frequency.

The signal  $x(t)$  is a function of time, and its STFT is a function of time and frequency. This transform is linear and depends on the chosen window,  $w$ . Details on how to choose the parameters of the observation window, as length and shape, and the discrete-time version, are presented in [20] [21] [22].

The Wavelet Transform (WT) was promoted to solve the time-frequency resolution problems of Fourier-type methods. A concept called "multi-resolution" or "multi-scale" is promoted. In the case of continuous time wavelet transform, a basis of translated and dilated functions called wavelets are used as

$$\psi_{t',a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-t'}{a}\right) \quad (3)$$

The wavelet transform is then

$$WT_{xw}(t, a) = \int_{-\infty}^{\infty} x(\tau)\frac{1}{\sqrt{a}}\psi^*\left(\frac{\tau-t'}{a}\right)d\tau \quad (4)$$

The Cohen class is the set of all bilinear representations, invariant under time and frequency translations, and described by the equation

$$\begin{aligned} C_x(t, f) &= TFR(x(t, f)) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(t_1, t_2; t, f)x(t_1)x^*(t_2)dt_1dt_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(t, f; v, \tau)x\left(v + \frac{\tau}{2}\right)x^*\left(v - \frac{\tau}{2}\right)dv d\tau \end{aligned} \quad (5)$$

with

$$k(t, f; v, \tau) = K\left(v + \frac{\tau}{2}, v - \frac{\tau}{2}; t, f\right) \quad (6)$$

where the kernel  $k(t, f; v, \tau)$  has some special properties, as discussed in [23]. By an equivalent parameterization the equation (5) becomes

$$\begin{aligned} C_x(t, f) &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{t-f}(t-v, f-v) \cdot W_x(u, v)dv d\tau \end{aligned} \quad (7)$$

with

$$\phi_{t-f}(t, \tau) = K(t, 0; 0, \tau) \quad (8)$$

and

$$W_x(t, f) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \quad (9)$$

The function  $W_x(t, f)$  is called the Wigner-Ville distribution (WVD), [20] [24], being one of the most important members of the Cohen's class. It may be the only distribution with real values that satisfies the properties necessary for the classical applications of signal processing. It is also the only distribution to provide perfect localization for impulse signals and signals with a linearly modulated frequency, [24]. In particular, the WVD is always real-valued; it preserves time and frequency shifts and satisfies the marginal properties. It has also some drawbacks, as the apparition of the cross-terms. This is the reason for using the Choi-Williams Distribution (CWD) [24], where the kernel function is

$$\phi(t, \tau) = \exp\left[-(t \cdot \tau)^2 / \sigma^2\right] \quad (10)$$

This distribution function adopts exponential kernel to suppress the cross-term that results from the components that differ in both time and frequency centers.

The discrete time WVD is defined by [24]

$$W(n, m) = \frac{1}{2N} \sum_{k=0}^{N-1} x(kT) \cdot x^*((n-k)T) \cdot \exp\left(-\frac{j\pi \cdot m \cdot (2k-n)}{N}\right) \quad (11)$$

It is informal to verify that  $W(n, m)$  is a periodic function of period  $2N$  in both time and frequency. The last relationship shows that in the range  $0 \leq n < 2N - 1, 0 \leq m < 2N - 1$ , representing one period, the WVD needs only be calculated over the range  $0 \leq n < N - 1, 0 \leq m < N - 1$ , having an area of one quarter that of the complete period.

The coefficients of the time-frequency transform define an image, which will be called a Time-Frequency Image (TFI).

### B. Entropy Transform

The Renyi entropies are important measures of the information, in wide sense. The measures are scale-dependent when applied to continuous distributions, so their absolute values are meaningless. Therefore, they can generally only be used in comparative or differentiable processes. The information content and the complexity of a probability density function can be measured by this entropy function. The Renyi entropy is intensively used in the field of statistical signal processing, especially in non-stationary conditions, being able to estimate the number of the

components of complex signals and the degree of randomness in various signal representation framework, in time or frequency domains [25] [26] [27].

In the case of continuous signal  $X(t)$ , the Renyi's entropy of the order  $\alpha$  is defined as:

$$H_\alpha(X) = -\frac{1}{1-\alpha} \log_2 \int f_X(x) dx, \alpha > 0, \alpha \neq 1 \quad (12)$$

where  $f_X(x)$  is the probability density function (pdf). For univariate discrete signals the common expression for the  $\alpha$ -order Renyi entropy is

$$H_\alpha(P) = -\frac{1}{1-\alpha} \log_2 \sum_{j=1}^N P_j^\alpha, \alpha > 0, \alpha \neq 1 \quad (13)$$

As relations (12) and (13) show, the computation of the entropies needs the availability of the exact or estimated pdf. (The probabilities' set, in discrete case). There are estimation solutions based on, e.g., Gaussian kernels, which provides expression as [28]

$$\hat{H}_\alpha(X, \sigma) = \frac{1}{1-\alpha} \log \left[ \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{N} \sum_{k=1}^N \left( G(X(n) - X(k), 2\sigma^2) \right) \right)^{\alpha-1} \right], \quad (14)$$

$\alpha > 0, \alpha \neq 1$

The entropy estimators require the selection of the kernel size,  $\sigma$ . This should be small (relative to the standard deviation of the data). Values between 0.1 and 2 for unit-variance signals are good choices, [29].

In the case of images,  $\mathbf{I}$ , a normalized image is considered as a probability density function. The  $\alpha$ -order Renyi entropy from [25] is considered as

$$HR_\alpha(\mathbf{I}) = -\frac{1}{1-\alpha} \log_2 \iint \left( \frac{I(t, f)}{\iint I(u, v) du dv} \right)^\alpha dt df \quad (15)$$

By discretization of this measure, i.e., by setting  $t=n \cdot \Delta t$  and  $f=k \cdot \Delta f, n, k \in Z$ , it results an expression as

$$HR_\alpha(\mathbf{I}) = -\frac{1}{1-\alpha} \log_2 \sum \sum \left( \frac{I[n, k]}{\sum \sum I[n', k']} \right)^\alpha + \log_2(\Delta t \cdot \Delta f) \quad (16)$$

By comparing the entropy of two images, associated to consecutive frames – as evolution in time, is possible to detect the differences and thus to make change detection.

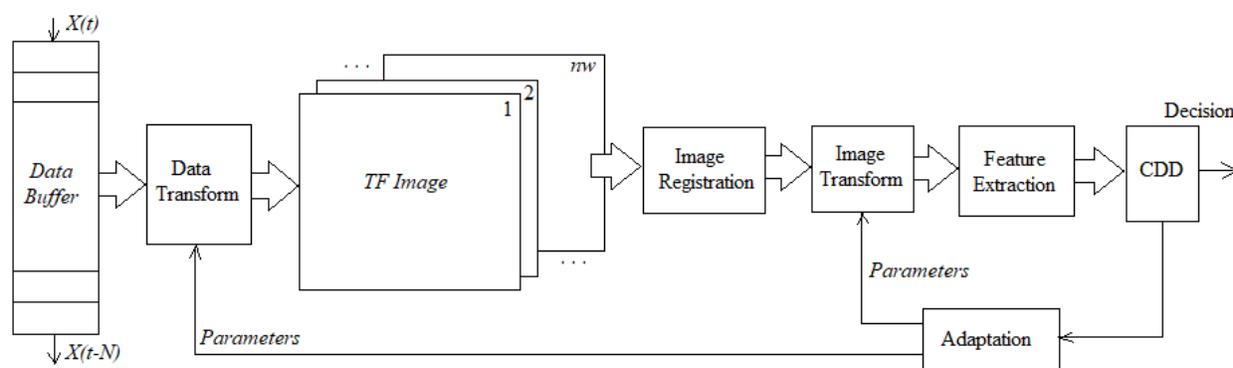


Figure 3. The general structure of the method for selection and extraction of features

### III. DESCRIPTION OF THE METHOD

Considering some results obtained by other previous studies and works, e.g., [13], a method to select and to extract the features of the time-frequency image and to define a new set of features, is developed.

The structure of the method is presented in Figure 3. Data coming from environment/process is stored in a data buffer for analysis and processing. Depending on objective, a data transform, e.g., Choi-Williams time-frequency transform, is applied to obtain an image. Depending on processing resources and time-constants of the analyzed signals and process, a set of TFIs is obtained. The set of these images is then registered based on the detection and computation of the main components in the analyzed image. The registered image is the base for feature selection process. The selected features could have physical meaning, as frequency, energy, bandwidth or spreading, or might be generated from other image transforms, as, e.g., Hough transform [30] [31] or Discrete Cosine Transform [32] [33]. In this work, a transform based on the representations of isolines (line of equal elevation) of a matrix is applied [34][35]. The features of these contours, as number, area, average, etc. could be used as feature for CDD objectives. Both data transforms, i.e., for time-frequency image and feature generation need parameters. The number and the values of these parameters depend on CDD performance/results. Thus, an adaptation block is necessary in order to establish the necessary parameters for each transform, as presented in Figure 3.

#### A. Vibration Data

Data were considered for the case of faults in bearings, available from [36], which are also well explained and analyzed in [37]. The number inside of the round parenthesis indicates the names of the files from the original source of data vibrations, i.e., [36]. Data are briefly described in Table I. Three types of faults are available, like F1 (Inner race), F2 (Ball) and F3 (Outer race). The case F0 means no faults. In the case of the fault F3, there are three sub-cases, depending on the fault position relative to the load zone: ‘centered’ (fault in the 6.00 o’clock position), ‘orthogonal’ (3.00 o’clock) and ‘opposite’ (12.00 o’clock), [37].

TABLE I. DATA TEST SET

Fault size	Faults in bearings			
	F1	F2	F3	F4
	Free	Inner Race	Ball	Outer Race
0.000 “	<b>d0</b> (97) (case#0)	-	-	-
0.007”	-	<b>d1</b> (105) (case#1)	<b>d2</b> (118) (case#2)	<b>d3</b> (130) (case#3)
0.014”	-	<b>d6</b> (169) (case#6)	<b>d7</b> (185) (case#7)	<b>d8</b> (197) (case#8)
0.021”	-	<b>d9</b> (209) (case#9)	<b>d10</b> (222) (case#10)	<b>d11</b> (234) (case#11)
0.028”	-	<b>d14</b> (3001) (case#14)	<b>d15</b> (3005) (case#15)	-

Vibration data from four sizes of the faults are available, data having the advantage of consistency, by considering faults from incipient/small size (0.007”) to larger (0.028”).

The sampling rate is 12,000 Hz, the motor is with no load, and all data are from drive end bearing (DE). A set of four classes of patterns are considered as: C#0 – no faults, defined by **d0**; C#1 –inner race faults, defined by {**d1**, **d6**, **d9**, **d14**}; C#2 – ball faults, defined by {**d2**, **d7**, **d10**, **d15**}; C#3 - outer race faults, defined by {**d3**, **d8**, **d11**}. The vectors **d4**, **d5**, **d12**, and **d13** correspond to other sites of the transducers. The set of classes are defined as

$$\text{Class \#1: } \mathbf{d0} // \text{Fault free} \quad (17)$$

$$\text{Class \#2: } \mathbf{d1, d6, d9, d14} // \text{Fault 1} \quad (18)$$

$$\text{Class \#3: } \mathbf{d2, d7, d10, d15} // \text{Fault 2} \quad (19)$$

$$\text{Class \#4: } \mathbf{d3, d8, d11} // \text{Fault 3} \quad (20)$$

For tests based on computer simulation, some new names for a variable were considered. All names beginning with “**d**” indicates a vector with 5,000 samples from normal conditions (no faults) and 5,000 samples from the records with faults. The variable **d0** contains the first 5,000 elements of the raw file named 97 from [36].

Figure 4 presents a sample of time varying signals from each considered class, by considering the first 1,000 samples from each file. The signals are scaled to [-1,1] by normalization. The waveforms seem to be quite different.

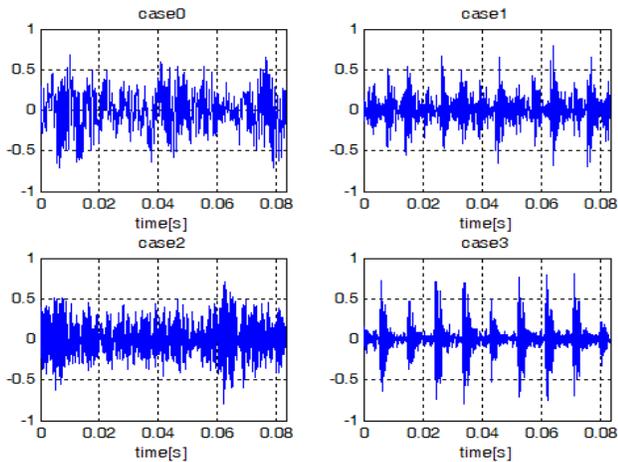


Figure 4. The structure for feature selection

**B. Time-Frequency Images (TFI)**

Figures 5-8 present a set of TFIs, one for each class. Frame or window no 5 is considered for all data records. There are also presented the time evolution, on the bottom side, from 0 to 0.08 [s], and the power spectral density, on the left side of each figure. For the case with no faults, i.e., Figure 5, the spectrum is centered roughly on 1,000 Hz. For the other cases, which cases with faults, the power spectrum density (psd) is spreading up to 4,000 Hz. The shape of the psds indicates some periodic components, like in the Figures 5-8, but also shapes close to the spectrum of modulated signals. These are signals with high frequency bandwidth, with spectral components from 500 Hz up to 4,000 Hz, like in Figure 6. These cases could generate real difficulties in processing and – later - in the detection and classification blocks.

In order to compute a prototype image for each class, the TFIs associated to data frames should be registered, and finally averaged to obtain a prototype image for each class. This is based on a stationarity hypothesis of the processed signals, and thus the main components from TFIs can be registered in both directions, horizontal and vertical. In order to keep the physical meaning/positions of these components, the registration techniques are restricted to horizontal translations only. The next subsection presents some aspects and results from the registration stage.

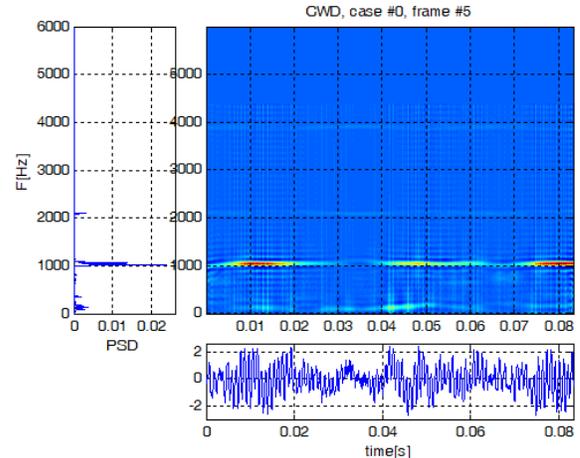


Figure 5. Time frequency image, Class #1 (free of faults)

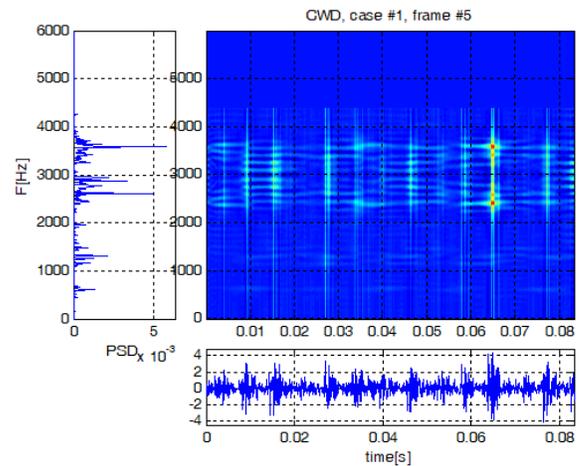


Figure 6. Time frequency image, Class #2

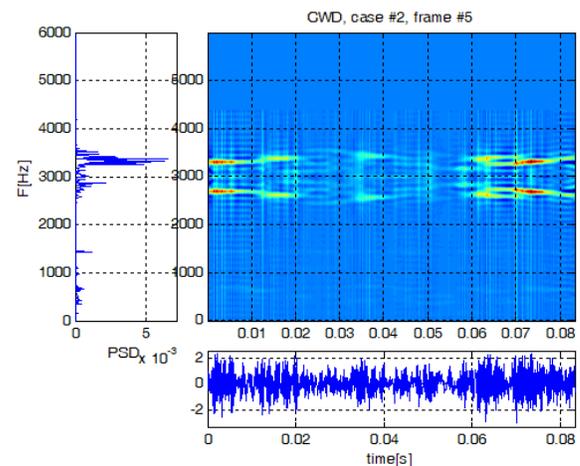


Figure 7. Time frequency image, Class #3

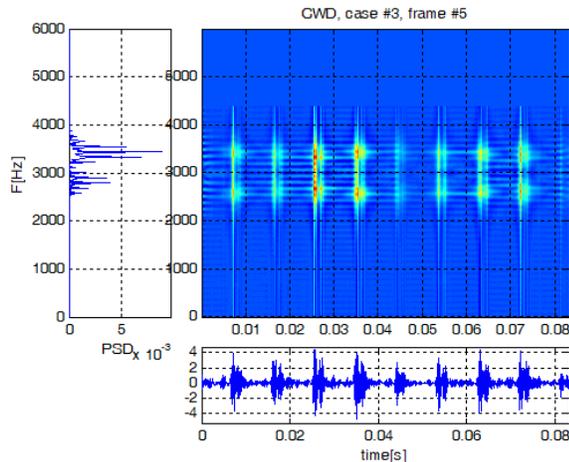


Figure 8. Time frequency image, Class #4

### C. Image registration

The registration process of the images is a pre-processing step, before the computation and the classification of the prototype images.

Common algorithms for image registration could use translation, on  $x$  and  $y$  directions; rigid processing, which means translation plus rotation; similarity, which means translation, rotation and scaling; affine transformation, which considers translation, rotation, scaling, and shearing. The choice of one of them is based mainly on the content of the image, the sources and the number of the images, which are considered for registration. Simple registration methods of the images, from the content point of view, use intensity-based registration algorithms. As complexity rises, the feature-based method is more indicated. Details and examples are available in many references as in, e.g., [38] [39].

The registration time is rapidly growing from translation to affine transformation. Sometimes, for complex transforms - like affine, the registration process could diverge. This is the reason to consider new methods valid for time-frequency images - in general - and in the case of bearings. The registration of TFI has only one degree of freedom, in the sense that any TFI processing must preserve the information of vertical axis, i.e., the frequency axis.

An adapted procedure considers several maxima from time-frequency image, which are considered as references, i.e., their positions remain unchanged during and after registration. Thus, the registration considers the physical meaning of these components, which should have the physical parameters (e.g., frequency), whatever the moment on the time axis. There should be also a distortion limit, in the sense that all images, which are far of the reference, should be removed from the registration set. More details are available in [40]. Further works might investigate some filtering techniques, in order to eliminate the noise or other components.

The reference and the registered images, in pair for each class, are presented in the Figure 9, for a window length of 2,000 samples.

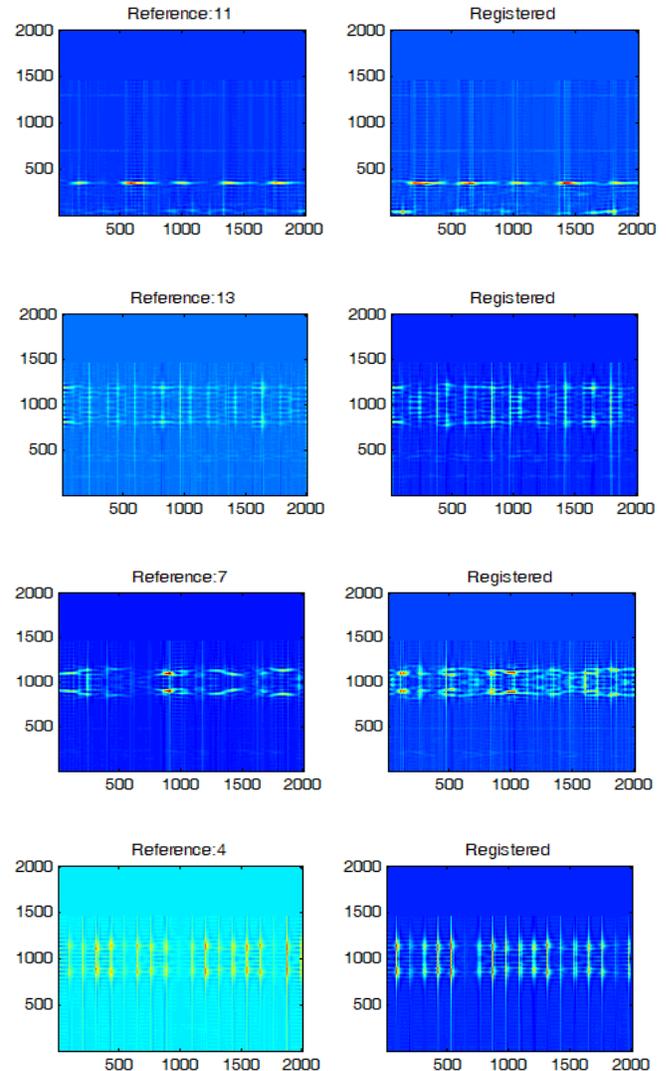


Figure 9. The reference and the registered images, for classes #1 to #4

### D. Feature selection

All images are quite complex, by having many components with various shapes. A first exploratory idea in order to describe the complexity of these images is to compute the Renyi entropies. The numeric results are presented below by the Table II with bold numbers and the averages values of the classes are represented in Figure 10, which does not show important differences among them - especially for the case of faults. As example, cases faults F2 and F4 are difficult to distinguish, the difference between entropies being 0.02 only. The range values of the Renyi entropies indicate the possibility of change detection but difficulties for classification.

An improvement in describing more accurate the content of the TFIs is to consider the shape of the main peaks from the analyzed image. A pre-defined number of peaks could be considered, e.g., 1 to 3, depending on the complexity of the image. Thus, a new image is considered and defined in terms of contours, defined by the above peaks, which will

be called transformed image or contour-based image (CBI). This transformation reduces the computation task, by keeping the information about the shape and position of the main components.

TABLE II. THE RENYI ENTROPIES OF THE REGISTERED IMAGES

	Faults / Classes			
	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
	<i>Free</i>	<i>Inner Race</i>	<i>Ball</i>	<i>Outer Race</i>
0.000 "	<b>4.40</b> (case#0)	-	-	-
0.007"	-	<b>6.08</b> (case#1)	<b>5.93</b> (case#2)	<b>5.64</b> (case#3)
0.014"	-	<b>4.25</b> (case#6)	<b>5.89</b> (case#7)	<b>6.18</b> (case#8)
0.021"	-	<b>4.60</b> (case#9)	<b>5.77</b> (case#10)	<b>4.09</b> (case#11)
0.028"	-	<b>6.24</b> (case#14)	<b>5.30</b> (case#15)	-
Average	<b>4.40</b>	<b>5.29</b>	<b>5.72</b>	<b>5.31</b>

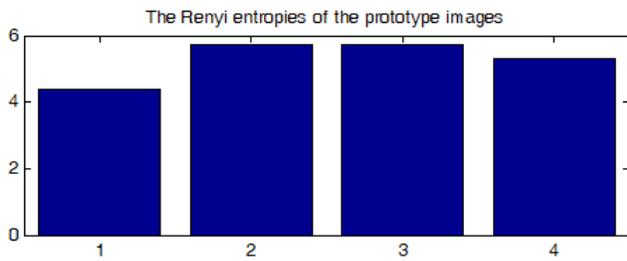


Figure 10. The averaged Renyi entropies of the registered images

In the set of the next two figures, i.e., Figure 11 and Figure 12, the raw/original images and the transformed images are presented, for classes C#1 and C#2. On blue background, the set of the registered images are presented. The registered images are considered the prototypes of the classes, from pattern recognition point of view.

The transformed images are presented on white background. Some details are presented in Figure 13, for three values of the number of contours ( $nc = 1, 2, \text{ and } 3$ ). As the number of contours is rising, the shape is coming more complex.

- a) the common content of the images is of vertical curves, as C#1(1,6,9), C#3(3,11,12);
- b) the class C#2 has a very complex pattern, for all cases (2,7,10, and 15);
- c) the classes C#1 and C#3 have some strange patterns, C#1(14) and C#3(8). Keeping all these images will damage the final classification.
- d) the vertical curves/contours of the transformed images are in fact a set of contours, which could be described each by numbers and areas. These could be used as features of the CBIs and later for the associated fault.

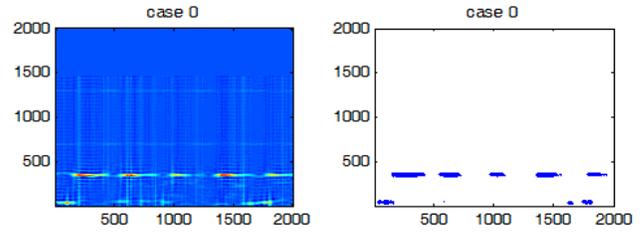


Figure 11. Original and transformed image, Class #1

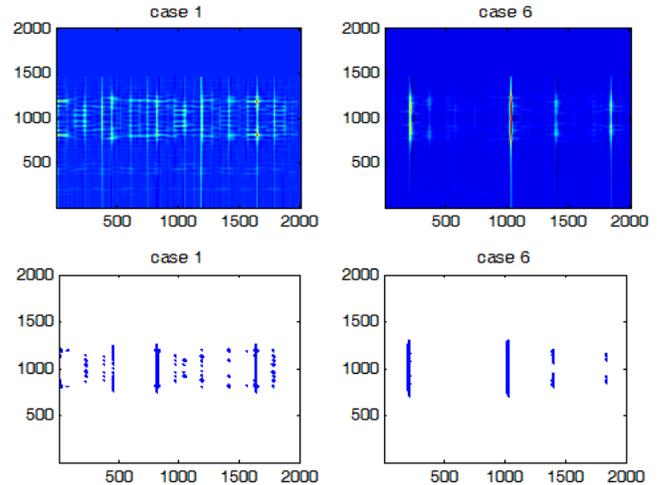


Figure 12. Original and transformed images, class #2 and #3.

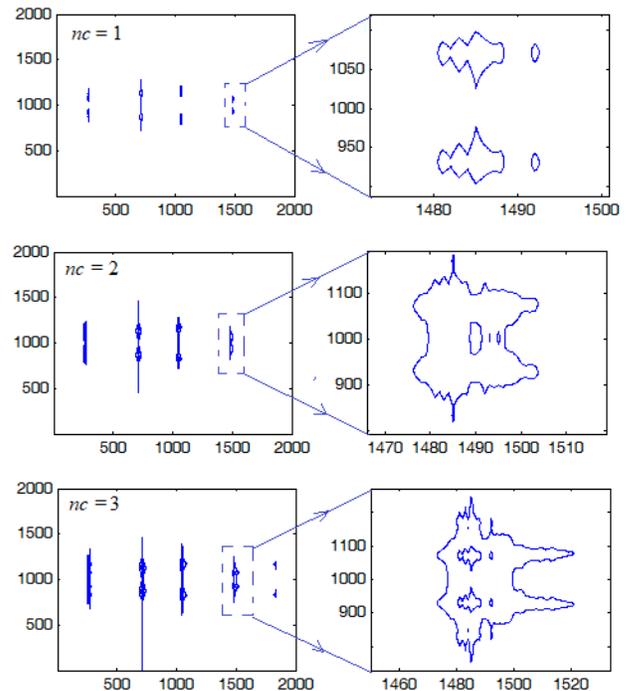


Figure 13. Details of contours based images,  $nc = 1, 2, 3$ .

In order to extract the right information, as features from the transformed images, some elements or parameters should be considered and properly used. In this work, parameters based on the number and size of the contours are used, mixed with the values of some statistical moments.

The basic elements of the feature vector are:

- i) the number of contours,  $N_c$ , as a measure of the complexity;
- ii) the area of the polygons,  $A_c$ , as a measure of the spreading on horizontal plane;
- iii) the variance of the above areas,  $var(A_c)$ , as a measure of the complexity;
- iv) the average of the area of the polygons,  $E\{A_c\}=A_c$ ;
- v) the mean of the squared values of areas,  $E\{A_c^2\}$ ;
- vi) the Renyi entropy of transformed images,  $RH$ .

A vector of features is defined for each class by using the above features, as

$$\mathbf{f}_i = \left[ N_c \quad \sum A_c \quad var(Ac) \quad \overline{A_c} \quad \overline{A_c^2} \quad RH \right], \quad (21)$$

$i = 1,2,3,4$

For each data vector,  $\mathbf{d}$ , from a class, the vector  $\mathbf{f}_i$  is evaluated, and matrix of features is obtained for each class, as

$$\mathbf{F}_j = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \dots \quad \mathbf{f}_4], \quad j = \overline{1,4} \quad (22)$$

The effect of the features is estimated by a general discriminant matrix of the classes

$$D(k, j) = \sum_{k=1}^4 \sum_{j=1}^4 (\mathbf{F}_k - \mathbf{F}_j)^2, \quad k, j = \overline{1,4} \quad (23)$$

or, by considering only the distinct classes, by the discriminant function

$$D_1 = E\{D(k, j) | k \neq j, k > j, k, j = \overline{1,4}\} \quad (24)$$

or

$$D_2 = \sum_{k=1}^4 \sum_{j=1, k \neq j}^4 D(k, j) \quad (25)$$

This should be high as possible. The number of the features considered in the Equation (21) could be modified, in order to gain the highest dissimilarity among classes.

#### IV. RESULTS OF THE EXPERIMENTS

The evaluation of the features for all classes is presented in Figure 14, with green for C#1, yellow for C#2, for C#3 and black for C#4. A high variance of the features for patterns of the same class is observed, e.g., feature 1 for class

#2, #3, and #4; feature 2 for classes #3 and #4; feature 6 for classes #2 and #3. If the variance is associated with the size of the fault, then a criterion to select the right features is to maximize the dissimilarity among classes.

The mean values of the features are presented in Figure 15, with different colors for classes. There six features for each class of four. There is a difficulty to make a good classification, especially for the classes 2, 3 and 4, where the evolution and the range of the values seem quite close. A solution to change this is to increase the number of contours ( $nc$ ) in CBIs from 1 to 2, and 3, with the results presented in Figure 16. Based on these evolutions, the next step is to evaluate a criterion for the selection of the best features for classification purposes, i.e., to maximize the dissimilarity of the classes in the feature space. Figure 17 shows the evolution of the dissimilarities in terms of the six features and based on Equation (23).

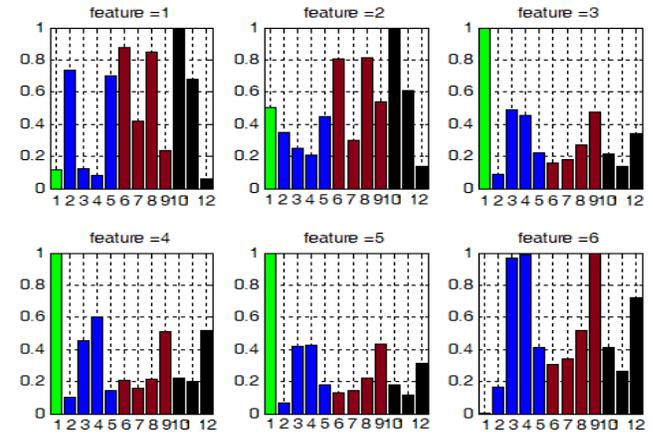


Figure 14. Feature space among various cases, from 1 to 12

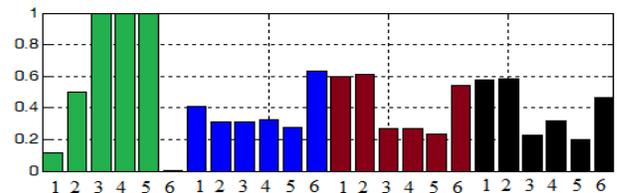


Figure 15. The mean values of the features;  $nc = 1$

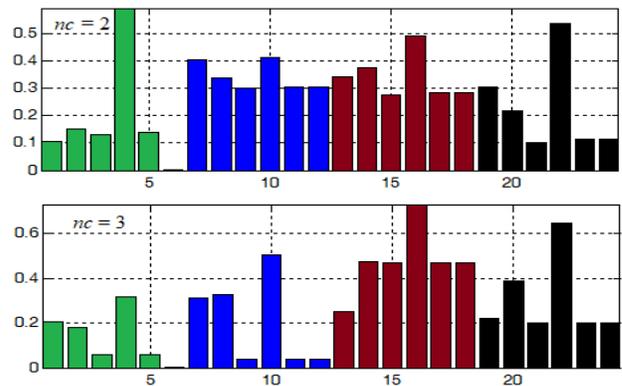


Figure 16. The mean values of the features;  $nc = 2$ , and 3

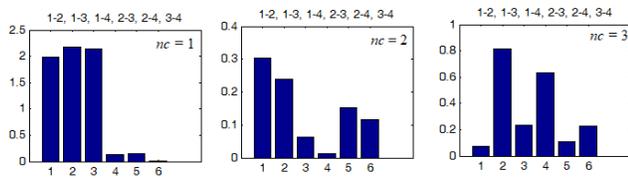


Figure 17. Dissimilarities among classes

The values of the discriminant functions (24) and (25) for various values of the number of contours are presented in Table III. The highest value of the discriminant functions is obtained for one contour only,  $nc = 1$ .

TABLE III. VALUES OF THE DISCRIMINANT FUNCTIONS

	$nc$				
	1	2	3	4	5
$D_1$	1.094	0.147	0.349	0.375	0.481
$D_2$	6.596	0.884	2.094	2.251	2.888

### CONCLUSION

The objective of the paper was to promote a method for feature selection and extraction from time-frequency images, as an alternative to some classical well-known methods, as those based on Hough Transform or Discrete Cosine Transforms.

Experiments used real vibration data coming from bearings of the rotating machines, bearings with various faults and sizes. The proposed method is general and can be applied also to other types of data, mechanically generated or not.

The roots of the method come from the fact that for classification purpose, the complexity of time-frequency images is not properly described by Renyi entropy. More information in terms of more features must be considered at the input of the classifier.

The method uses two data transforms. The first one is based on Choi-Williams time-frequency transform and the second uses a representation based on isolines of a matrix, applied to the main components of the time-frequency images. Before extracting the features, the time-frequency images are registered. Depending on the number of the contours obtained, which could vary, e.g., from 1 to 5, the features are varying and change the dissimilarity of the classes.

An important step is adaptation of the parameters for the used data transform. Depending on the dynamic properties of the process which generates the mechanical vibrations and depending on the evolution of the faults, the user must check the length of the data frame/window, the number of frames, the numbers of the contours which define the transformed images (contour – based).

The selected features are based on the statistical moments, as average, variance and squared average values of the areas of the contours. The information- based features is also used, by considering the Renyi entropy. Larger feature vectors could be considered by including also the number of the main components and their centers, in time and space.

The results show a good separability in the feature space, in the sense of clustering, and thus the possibility to obtain efficient classifiers.

The method could be extended to more complex signals and applications. The feature vector could be extended with qualitative or quantitative parameters, which describe the shape of contours, i.e., the distance to some standards cures as circles, ellipses or squares. Further research could consider also fuzzy logic in the description and selection of the features from time-frequency images.

### ACKNOWLEDGMENT

The work was partly supported by the Romanian Council for Research (UEFISCDI) under Grant 224/2014, “Experimental model for change detection and diagnosis of vibrational processes using advanced measuring and analysis techniques model-based (VIBROCHANGE)”.

### REFERENCES

- [1] D. Aiordachioaie, Th. D. Popescu, and B. Dumitrascu, “A Method of Feature Extraction from Time-Frequency Images of Vibration Signals in Faulty Bearings for Classification Purposes”, Proc. of International Conference on Emerging Networks and Systems Intelligence (EMERGING '2019), 22-26 September, 2019, Porto, Portugal.
- [2] J.H. Williams, A. Davies, and P. R. Drake (Eds), Condition-based Maintenance and Machine Diagnostics, Springer, 1994.
- [3] L. Fedele, Methodologies and Techniques for Advanced Maintenance, Springer, 2011.
- [4] M. Bengtsson, E. Olsson, P. Funk, and M. Jackson, “Technical Design of Condition Based Maintenance System”, Proceedings of the 8th Congress, University of Tennessee – Maintenance and Reliability Center, Knoxville, USA, May 2nd – 5th, Paper III, 2004.
- [5] K. Shin and J. K. Hammond, Fundamental of Signal Processing for Sound and Vibration Eng., John Wiley & Sons Ltd, 2008.
- [6] D. M. Yang, A. F. Stronach, and P. MacConnell, “The Application of Advanced Signal Processing Techniques to Induction Motor Bearing Condition Diagnosis”, P. Meccanica, Springer, 2003, vol 38(2), pp 297-308.
- [7] D. Mironovs and A. Mironov, “Vibration based signal processing algorithm for modal characteristics change assessment”, AIP Conference Proceedings 2029, 020043, 2018, pp.1-10.
- [8] B. Boashash (Ed), Time-Frequency Signal Analysis and Processing, Elsevier, 2016.
- [9] L. Stankovic, M. Dakovic, and T. Thayaparan, Time-Frequency Signal Analysis with Applications, Artech House, 2014.
- [10] M. Boufenar, S. Rechak, and M. Rezig, “Time-Frequency Analysis Techniques Review and Their Application on Roller Bearings Prognostics”. In: Fakhfakh T., Bartelmus W., Chaari F., Zimroz R., Haddar M. (Eds) Condition Monitoring of Machinery in Non-Stationary Operations. Springer, Berlin, Heidelberg, 2012.
- [11] B. Boashash, L. Boubchir, and G. Azemi, “A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals”, EURASIP Journal on Advances in Signal Processing, Article 117, 2012.
- [12] D. Rossetti, Y. Zhang, S. Squartini, and S. Collura, “Classification of bearing faults through time-frequency

- analysis and image processing," 2016 17th International Conference on Mechatronics - Mechatronika (ME), Prague, 2016, pp. 1-7.
- [13] D. Aiordachioaie, N. Nistor, and M. Andrei, "Change Detection in Time-Frequency Images by Feature Processing in Compressed Spaces", IEEE 24th Int. Symp. for Design and Technology in Electronic Packaging (SIITME), Iași, Romania, 2018, pp.181-186 .
- [14] E. Sejdic, I. Djurovic, and J. Jiang, "Time-frequency feature representation using energy concentration: An overview of recent advances", Digital Signal Processing, 19, 2009, pp. 153-183.
- [15] Q. Zhu, Y. Wang, and G. Shen, "Comparison and Application of Time-Frequency Analysis Methods for Nonstationary Signal Processing". In: Lin S., Huang X. (eds) Advanced Research on Computer Education, Simulation and Modeling. CISM 2011. Communications in Computer and Information Science, vol 175. Springer, Berlin, Heidelberg, 2011.
- [16] F. Hlawatsch and F. Auger, "Time-Frequency Analysis. Concepts and methods", ISTE Ltd and John Wiley & Sons, 2008.
- [17] M. Karjalainen and V. Pulkki, Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics, John Wiley & Sons, 2015.
- [18] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques", Mechanical Systems and Signal Processing, vol. 25 (6), 2011, pp. 2083-2101.
- [19] M. Akay, Time Frequency and Wavelets in Biomedical Signal Processing, Wiley-IEEE Press, 1997.
- [20] B. Boashash, "Introduction to Time-Frequency Signal Analysis". 10.1007/978-1-4612-0137-3\_11, 1997.
- [21] L. Cohen, "Time-Frequency Analysis", Prentice Hall, 1995.
- [22] J. Ville, "Theorie et applications de la notion de signal analytique", Câbles et transmissions, vol. 2A,1948, pp. 66-74.
- [23] H. Choi and J.W. Williams., "Improved time-frequency representation of multicomponent signals using exponential kernels", IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 6, 1989, pp. 862-871.
- [24] P. D. McFadden and W. Wang, Time-Frequency Domain Analysis of Vibration Signals for Machinery Diagnostics. (I) Introduction to the Wigner-Ville Distribution, University of Oxford, Report OUEL 1859/92, 1990.
- [25] P.A. Bromiley, N.A. Thacker, and E. Bouhova, "Shannon Entropy, Renyi Entropy, and Information", Statistics and Information Series (2004-004), 2004.
- [26] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring Time-Frequency Information Content Using the Rényi Entropies", IEEE Transactions on Information Theory, vol. 47(4), 2001, pp. 1391-1409.
- [27] D. Aiordachioaie, "Some properties of Entropies for Change Detection", The Annals Of "Dunărea De Jos" University Of Galati Fascicle III, 2013, Vol. 36, No. 1, ISSN 1221-454x Electrotechnics, Electronics, Automatic Control, Informatics, 2013, pp.43-48.
- [28] D. Erdogmus K.E.Hild II, and J.C. Principe, "Blind source separation using Renyi's alpha-marginal entropies", Neurocomputing, vol. 49 (1), 2002, pp. 25-38.
- [29] K.E. Hild II, D. Erdogmus, J.C. Principe, "An analysis of entropy estimators for blind source separation", Signal Processing, vol 86 (1), 2006, pp.182-194.
- [30] V.F. Leavers, Shape Detection in Computer Vision Using Hough Transform, Springer, 1992.
- [31] A. M. Tekalp, Hough Transform Methods, in Handbook of Image and Video Processing (Second Edition), Al Bovik Editor, AP, 2005.
- [32] K.R. Rao and P. Yip, Discrete Cosine Transform. Algorithms, Advantages, Applications, AP, 1990.
- [33] C. Loeffler, A. Ligtenberg, and G. Moschytz, "Practical Fast 1-D DCT Algorithms with 11 Multiplications", Proc. Int'l. Conf. on Acoustics, Speech, and Signal Processing, 1989 (ICASSP '89), pp. 988-991.
- [34] D. Hearn and M.P. Baker, Computer Graphics, Prentice-Hall, Inc., 1986.
- [35] M. Trott, The Mathematica GuideBook for Graphics, Springer-Verlag New York, 2004.
- [36] Case Western Reserve University, Bearing Data Center, At: <http://csegroups.case.edu/bearingdatacenter/home> , Retrieved: 06, 2019.
- [37] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study", Mechanical Systems and Signal Processing, 64-65, pp. 100-131, 2015
- [38] G. A. Ardeshir, Image Registration Principles, Tools and Methods, Springer, 2012.
- [39] I. N. Bankman (Ed.), Handbook of Medical Image Processing and Analysis, Elsevier, 2009.
- [40] D. Aiordachioaie and S. M. Pavel, "Qualitative Analysis of the Time-Frequency Images of Vibrations in Faulty Bearings", IEEE 11th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI), Romania, 2019, pp. 1-6.