An Automated Thermographic Image Segmentation Method for Induction Motor Fault Diagnosis

Petros Karvelis, George Georgoulas,
Chsysostomos D. Stylios
Laboratory of Knowledge and Intelligent Computing,
Technological Educational Institute of Epirus, Department
of Computer Engineering
Arta, Greece
{pkarvelis, georgoul, stylios}@kic.teiep.gr

Ioannis P. Tsoumas
Larges Drives Products R&D Department
Siemens Industry Sector - Drive Technologies
Nuremberg, Germany
ytsoumas@ieee.com

Jose Alfonso Antonino-Daviu, María José Picazo
Ródenas
Instituto de Ingeniería Energética
Universitat Politècnica de València
Valencia, Spain
joanda@die.upv.es

Vicente Climente-Alarcón
Department of Electrical Engineering and Automation
Aalto University, Espoo, Finland
viclial@ieee.org

Abstract—**Eventual failures in induction machines may lead to catastrophic consequences in terms of economic costs for the companies. The development of reliable systems for fault detection that enable to diagnose a wide range of faults is a motivation of many researchers worldwide. In this context, non-invasive condition monitoring strategies have drawn special attention since they do not require interfering with the operation process of the machine. Though the analysis of the motor currents has proven to be a reliable, non-invasive methodology to detect some of the faults (especially when assessing the rotor condition), it lacks reliability for the diagnosis of other faults (e.g. bearing faults). The infrared thermography has proven to be an excellent, non-invasive tool that can complement the diagnosis reached with the motor current analysis, especially for some specific faults. However, there are still some pending issues regarding its application to induction motor faults diagnosis, such as the lack of automation or the extraction of reliable fault indicators based on the infrared data. This paper proposes a methodology that intends to provide a solution to the first issue: a method based on image segmentation is employed to detect several failures in an automated way. Four specific faults are analyzed: bearing faults, fan failures, rotor bar breakages and stator unbalance. The results show the potential of the technique to automatically identify the fault present in the machine.**

Keywords—*Induction motor, fault diagnosis, SIFT, image segmentation, object matching*

I. INTRODUCTION

Induction motors are probably the most reliable rotating electrical machines [1]. Nonetheless, their participation in a vast number of industrial processes and applications, some of them critical, justifies the importance of developing suitable maintenance programs adapted to these machines. Within the induction motor predictive maintenance area, several techniques have been developed over recent decades that are intended to diagnose a wide range of eventual failures in these machines, when faults are still in their early stages of development. Emphasis has been put in the non-invasive nature of the developed diagnosis methods, since in many applications no interference with the machine operation is desired.

In this context, the techniques relying on the analysis of currents have had a significant prominence. The current is a quantity easy to be measured in a non-invasive way and the software and equipment required for its processing is also simple. This is the reason why techniques relying on current, from the classical Motor Current Signature Analysis (MCSA) [1] to the novel methods based on transient analysis of currents [2]-[6] have been extensively applied. However, in spite of the inherent advantages of current analysis, it is far from being the 'panacea' to completely determine the condition of the motor; despite it working well for detecting some specific faults (mainly those related to the rotor [1], [7]), it shows some constraints when trying to diagnose other faults such as bearing damages or insulation degradation, at least when these faults are in their early stages of development [7].

Infrared thermography is a non-invasive condition monitoring alternative that can provide very interesting information that can complement the diagnostics reached with other quantities. It has been extensively employed in the area of electric installations maintenance [8] and predictive maintenance of static electrical machines [9]. The few papers in the induction motors condition monitoring field, have focused on the diagnosis of faults such as: insulation failures in the magnetic circuit, and deficient connections or misalignments [10]-[14]. However, most of the works aim to detect simple failures, often external to the machine. Some recent works have proposed its extension to the detection of other faults [15], [16] showing interesting conclusions. These works prove that the

images obtained with infrared thermography can be helpful to detect hot spots or areas that can provide an alarm of an incorrect operation of the motor or any of its constituting elements [15]. This can motivate the user to carry out deeper analyses with other quantities. Also the profiles heating curves obtained at different points of the frame, change depending on the fault present in the machine, a fact that can be also informative about its condition.

Though the cost of the infrared cameras is still high, its use as a complementary tool for induction motor condition monitoring can be justified in large critical motors where the investment in such cameras is compensated by the higher reliability that can be achieved in the diagnostic and that can avoid significant losses in terms of unplanned production shutdowns.

However, in spite of the potential of the methodology, its integration in currently existing monitoring systems is still a pending issue. The lack of automation of the methodology is a reason for that: a user expertise is needed to identify an anomaly in the temperature distribution of a certain region that is linked with a certain fault. This makes its implementation in on-line systems or condition monitoring devices difficult.

This paper proposes a method to automate the pattern recognition process in infrared images by object matching. Object matching in two-dimensional (2-D) images has been an important topic in computer vision, object recognition and image analysis [17], [18], [19].

The proposed method is applied in this work for the detection of four possible faults: bearing faults, fan failures, rotor bar breakages and stator unbalances. The results show the ability of the technique to correctly identify the fault present in the machine by simply analyzing the infrared images obtained from experiments under the aforementioned faulty conditions. The rest of the paper is structured as follows: Section II summarizes all the involved techniques as well as the experimental set up. In Section III the validation protocol is described and the achieved results are presented. Finally Section IV concludes the paper offering also some insights for future research.

II. METHOD

For any object in an image, interest points on the object can be extracted to provide a "feature description" of the object. The problem can be modeled as following: having an input image (Train Image) displaying the object of interest, we try to find and match the object against a number of images (Target Images). These methods first try to extract a number of points of interest from the Train Image and the Target Image. Then they try to match these points using a geometric and photometric invariant descriptor. In our case we are not only interested in the detection of the matching of the induction machine using Train and Test images but also discriminating the temperatures of the machine into three parts. Thus prior to the application of our method we have segmented the induction machine image into three parts using an image of a reference machine. The different parts of the machine compose the outline of the machine and define the model image $M(x, y)$ consisting of the following parts (Fig. 1):

- Part A: coupling.

- Part B: frame.
- Part C: fan cowl.

Our method is composed by a number of steps: a) extract the Scale Invariant Feature Transform (SIFT) [20], [21] keypoints from the Train image, b) extract the SIFT keypoints from the Target image, c) match the SIFT points from the two images computing the geometric transformation between the two images, d) match the model from the Train Image to the Test image using the geometric transformation computed previously, d) extract the mean intensity of each part of the machine as well as the surrounding area and e) feed these temperatures to a classifier in order to diagnose the type of fault. The steps of our method are presented in the Fig. 2.
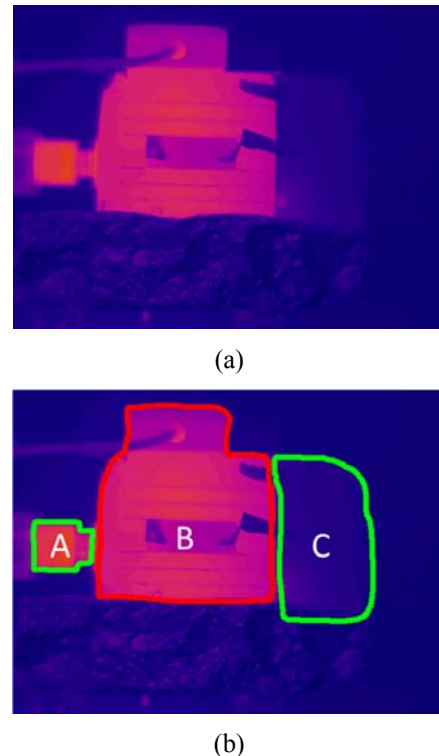

(a)

(b)

Fig. 1.   The three parts of the induction machine. a) the reference machine, b) the three parts depicted on the machine image.

A. Image Acquisition

Several experiments were carried out by using a 1.1 kW induction motor (reference 1LA2080-4AA10). In the tests, the machine was started direct on-line. The motor was driving an auxiliary DC machine acting as a load. A power analyzer was employed to measure the electric power, whereas a torque transducer and a digital tachometer were providing, respectively, the mechanical torque and the speed. The test-bed is displayed in Fig. 3. An infrared camera was employed to capture the images of the motor. The camera was a high-sensitivity long wave FLIR S65 Series, with a firewire connection enabling its interfacing with a portable computer. The computer was fitted with an acquisition and analysis software that permitted the visualization of the captured images. This software enabled knowing, with high accuracy, the temperature distribution on the motor frame at any time.
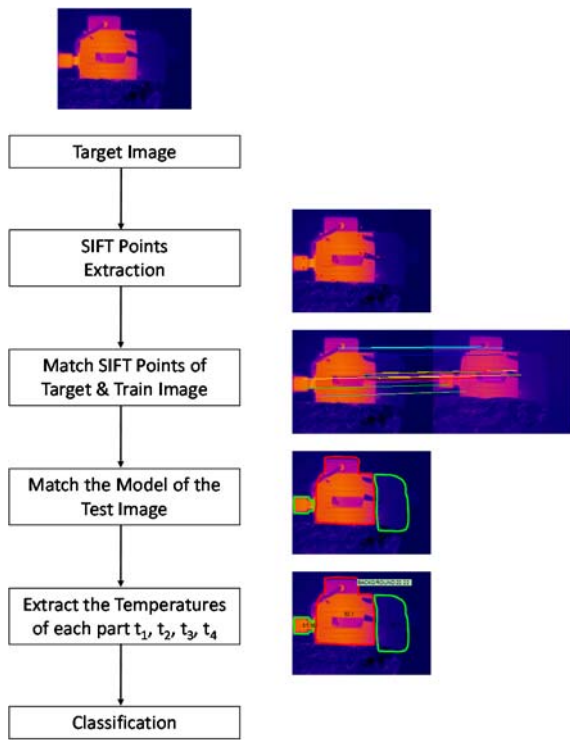
Fig. 2.   The steps of our method.

Fig. 3.   Experimental test bed

B. Image Segmentation

1) SIFT Points extraction

One of the most useful methods in order to define interest points is Lowe's method [20], [21] or best known as SIFT. SIFT generates a large collection of feature vectors from the image each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion.

The computation of SIFT descriptors is divided into four major stages: (a) scale-space extrema detection (b) keypoint localization; (c) orientation computation; and (d) descriptor computation.

a) Extrema Detection

In this stage the location of each SIFT keypoint is detected using the difference of Gaussian functions $D(x, y, \sigma)$ which is computed as the difference of two images, one with k scales times the other:

$$D(x,y,\sigma) = G(x,y,\kappa\sigma) * I(x,y) - G(x,y,\sigma) * I(x,y) \text{ ,(1)}$$

where I the initial Image, σ is the scale or the standard deviation of the Gaussian function used in the convolution ($*$) with the initial image:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} . \tag{2}$$

b) Interest Point Localization

For this stage the SIFT algorithm eliminates unstable interest points by finding those that have low contrast or are poorly localized on an edge. In order to achieve this, the Laplacian value for each interest point is computed. The location of extremum, is given by the following equation:

$$\mu = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} . \tag{3}$$

c) Orientation Computation

For this stage we assign an orientation to each keypoint by the computation of the gradient vector with magnitude m and angle a of each point. This is achieved easily using pixel differences:

$$m = \sqrt{\left(I(x,y) - I(x+1,y)\right)^2 + \left(I(x,y) - I(x,y+1)\right)^2} \text{ ,(4)}$$

$$a = tan^{-1}\left[\frac{I(x,y) - I(x,y+1)}{I(x,y) - I(x+1,y)}\right] . \tag{5}$$

d) Descriptor computation

Finally the SIFT descriptor is a weighted histogram of the gradient orientations and locations in a neighborhood surrounding the interest point [21].

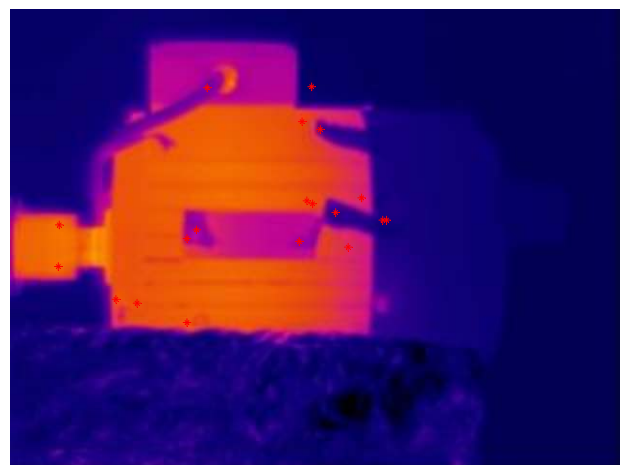The detection of a number of SIFT points for a test image is shown in Fig. 4.

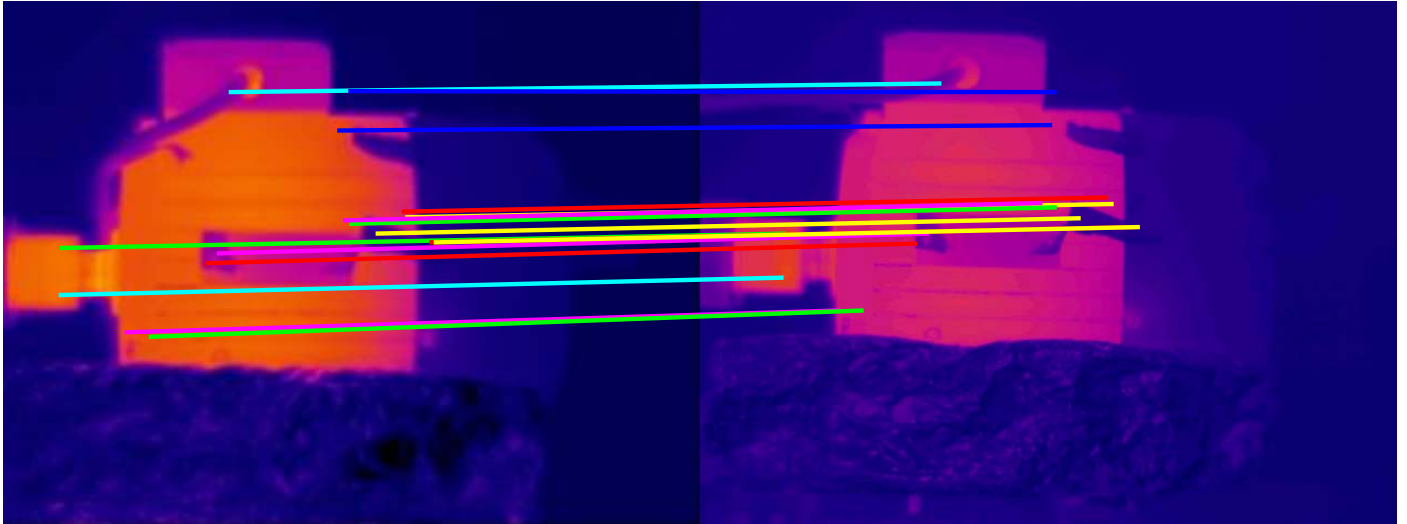Fig. 4.   SIFT points depicted as red stars on the themrographic image

3398

Fig. 5. Mathing of the SIFT points between .the Target Image (on the left) and the Train Image (on the right).

2) Match the SIFT Points of the Target and the Train Image

The set of image descriptors between the Train and Target image can be mutually matched for each keypoint by finding the keypoint in the other image domain that minimizes the Euclidean distance between the descriptors. In order to suppress some ambiguous matches Lowe [20] only accepted matches for which the ratio between the distances to the nearest and the next nearest points is less than 0.8. A speed up of the previous algorithm was proposed also by Lowe [21] based on best-bin-first (BBF) algorithm [22]. An example of the matching of the SIFT keypoints between the Target and the Train image is shown in Fig. 5.

Having matched the SIFT points of the Target and the Train image we also use our train model image to match it to the target image. This is shown in the following Fig. 6 were the final perimeter of the model is depicted in the Test image.
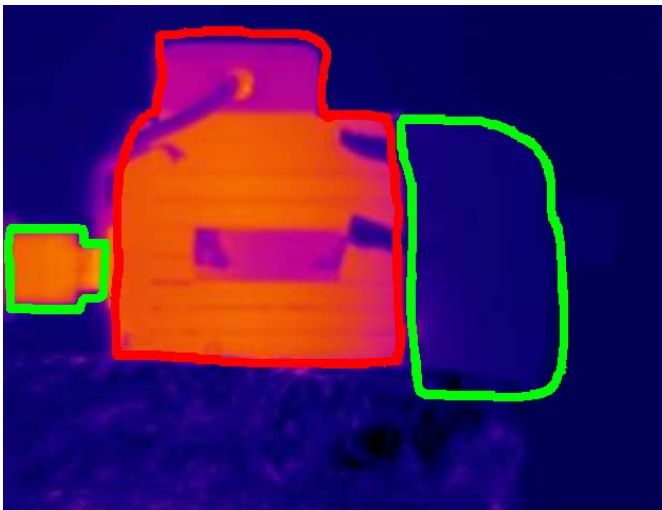
3) Extract the Temperatures of each part

In order to extract the temperatures ($\bar{t_i}$, $i = 1,2,3$) of each region of the machine (R_1, R_2, R_3) displayed on the thermographic image, we extract the perimeter of the model image (Fig. 7 (a)).

$$P(\text{x}, \text{y}) = \begin{cases} 1 & if\ M(\text{x}, \text{y}) = red\ or\ green \\ 0 & otherwise \end{cases}. \qquad (6)$$

Using this image as a mask we fill the three areas [23] producing three segmented areas described by the binary image $B(\text{x}, \text{y})$ (Fig. 7 (b)).

$$B(\text{x}, \text{y}) = \begin{cases} 1, & if\ P(x, y) \in R_i, i = 1,2,3 \\ 0, & if\ P(x, y) \in R_0 \end{cases}. \qquad (7)$$
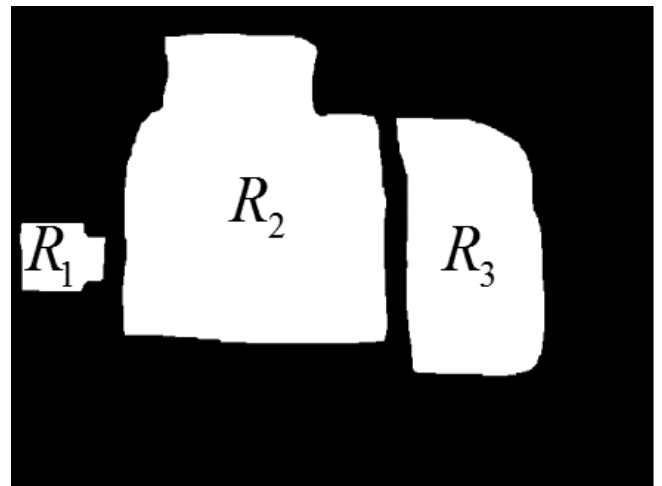

Fig. 7. The three regions of interest segmented from the Target image. a) The perimeter $P(\text{x}, \text{y})$ of the three parts of the model image, and b) the three segmented regions.

Fig. 6. The model image (describing the three parts of the induction machine) matched in the Target image.

Having segmented the three areas of the image we compute the average temperature $\bar{t}_i$, $i = 1, 2, 3$ of each area by computing the average intensity of each region of the thermographic image.

$$\bar{t}_i = \frac{1}{|R_i|} \sum_{p \in R_i} I(p), \ i = 0, 1, 2, 3 , \quad (8)$$

where $|R_i|$ the number of pixels of the region R_i, $i = 1, 2, 3$.

In the next step we compute the ratio of each part temperature to each other part temperature to be used as features to the next stage. This is shown in Table I.

TABLE I.   THE RATIOS OF THE MEAN TEMPERATURES OF EACH REGION

Ratio 1 ($Ratio_1$)	Ratio 2 ($Ratio_2$)	Ratio 3 ($Ratio_3$)
$\dfrac{\bar{t}_1}{\bar{t}_2}$	$\dfrac{\bar{t}_2}{\bar{t}_3}$	$\dfrac{\bar{t}_3}{\bar{t}_1}$

C. Classification

1) C4.5 Tree Classifier

For our classification scheme we use a tree classifier namely the C4.5 [24], [25]. Decision trees are among the most powerful approaches in knowledge discovery and data mining. The C4.5 algorithm constructs a decision tree with a divide and conquer strategy. For each node in the tree, a number of a set of cases is associated. Cases are also assigned weights to take into account unknown attribute values.

Let T be the cases associated at the node t then we compute the weighted frequency $f(C_i, T)$ of cases in T whose class is $C_i, i \in [1, N_C]$ where N_C the number of classes. If T contains cases belonging to two or more classes then the information gain must be computed for each attribute:

$$gain = \text{info}(T) - \sum_i \frac{|T_i|}{|T|} \times \text{info}(T_i) , \quad (9)$$

where $T_i, i = 1, ..., s$ is the set of the splitting produced by the test on the selected attribute and s is the number of splitting of node t. Finally $\text{info}(T)$ is the entropy function:

$$\text{info}(T) = \sum_{j=1}^{N_C} \frac{f(C_j, T)}{|T|} \times \log_2 \left(\frac{f(C_j, T)}{|T|} \right). \quad (10)$$

If T_i is not empty, the divide and conquer approach consists of recursively applying the same operations on the set consisting of T_i plus those cases in T with unknown value of the selected attribute. The algorithm in pseudo code is shown below [30]:

```
Algorithm C4.5(D)
Input: an attribute-valued dataset D
Tree = {}
if D is "pure" OR other stopping
criteria met then
    terminate
end if
for all attribute a ∈ D do
    Compute information-theoretic
    criteria if we split a
end for
a_best = Create a Decision node to above
computed criteria
Tree = Create a decision node that tests
a_best in the root
D_u = Induced sub-datasets from D based
on a_best
for all D_u do
    Tree_u=C4.5(D_u)
    Attach Tree_u to the corresponding
    branch of Tree
end for
return Tree
```

2) Naïve Bayes Classifier

The Naive Bayes classifier [26] is based on Bayes' theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful.

Let $Z = \{z_1, z_2, ..., z_N\}$ defining the set of temperatures for the image dataset where each $z_i \in \Re^3, i = 1, ..., N$ described the mean temperatures of the regions of the image i (defined in Table I), and N is the number of images. Using Bayes' rule and assuming that the prior probabilities are equal for each class namely C_1, C_2, C_3, C_4, C_5 we can compute the class of an unknown sample z_i as:

$$C^* = \underset{1 \le n \le 5}{arg\ max}\ p(z \mid C_n). \quad (11)$$

In order to estimate the likelihood $p(z \mid C_n)$ the Naïve Bayes assumption defines that the features are conditionally independent given the category, i.e.:

$$p(z_i \mid C_n) = \prod_{j=1}^{3} p(z_i^j \mid C_n). \quad (12)$$

where z_i^j is the temperature of the $j - th$ region of the $i - th$ image.

III. RESULTS

A 10-fold cross validation procedure was employed in order to access the performance of our method for each classifier [27]. The classification results for each classifier are presented in detail below. Both classification algorithms were implemented using WEKA [26], [27]. In Fig. 8 below we display the extracted features (ratios of Table I) of each image of our dataset into a three dimensional space using a different color for each type of fault (Rotor Bar Breakage, Bearing Faults, Stator Unbalance, Fan Failures) and also including the healthy class.
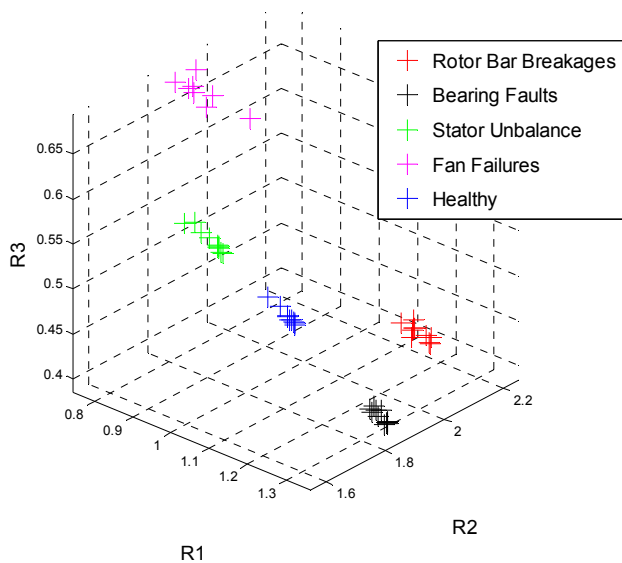
Fig. 8. The ratios displayed into a three dimensional space.

1) C4.5 Tree and Decision rules

For the C4.5 classification scheme the overall classification accuracy is 91.48%. Table II, presents the confusion matrix of the C4.5 classifier.

TABLE II. C4.5 CONFUSION MATRIX

True Class	Predicted Class				
	Healthy	*Bearing*	*Unbalance*	*Fan*	*Bars*
Healthy	9	0	0	0	0
Bearing	1	9	0	0	0
Unbalance	1	0	9	0	0
Fan	1	0	0	7	0
Bars	0	1	0	0	9

The classification rules from the C4.5 classification algorithm are described below:

```
1. If ( Ratio₃ <=0.41 && Ratio₁ <=1.14)
   class=BARS

2. If( Ratio₃ <=0.41 && Ratio₁ >1.14)
   class=BEARING

3. If( Ratio₃ >0.41  && Ratio₂ <=1.65)
   class=UNBALANCE

4. If( Ratio₃ >0.41  && Ratio₂ >1.65 && Ratio₁
   <=0.80)
   class=FAN

5. If( Ratio₃ >0.41  && Ratio₂ >1.65 && Ratio₁
   >0.80)
   class=HEALTHY
```

2) Naïve Bayes.

For the Naïve Bayes classification scheme the overall classification accuracy was 100%. Table III, presents the confusion matrix of the Naïve Bayes classifier.

TABLE III. NAÏVE BAYES CONFUSION MATRIX

True Class	Predicted Class				
	Healthy	*Bearing*	*Unbalance*	*Fan*	*Bars*
Healthy	10	0	0	0	0
Bearing	0	10	0	0	0
Unbalance	0	0	10	0	0
Fan	0	0	0	8	0
Bars	0	0	0	0	9

IV. CONCLUSIONS

In this work we presented an automated segmentation method of thermographic images which can be used as part of a fault diagnosis method for induction machines. The invariance of the method to image translation, scaling, and rotation makes it perfect for its application to industrial settings where having a camera fixed to a specific position is almost impossible. Therefore the camera can be used as a portable monitoring device followed with the appropriate processing steps that were described in this study.

Our approach however requires the development of a model for the induction machine under study which requires some effort. On the other hand once such a model is defined then the whole procedure can easily be used provided that the necessary training data are available (something that applies to all data driven methods).

In this study we used two of the most popular data mining algorithms. If someone is willing to sacrifice a bit of accuracy for interpretability of results, C4.5 is the method of choice. On the other hand Naïve Bayes even though considered a less powerful approach, under the specific setting managed to have perfect discrimination. Due to the restricted number of examples no definite conclusions can be drawn. Nevertheless the scatterplot of relative temperatures reveals that each fault occupies a different area in this figure despite the fact that not all images were acquired from the exact same point (however their boundaries are not so easily described by planes parallel to the axis (a characteristic of decision trees boundaries) and this could be an explanation of the inferior results of the C4.5 along with its inherent pruning mechanism which in combination with the restricted number of cases prevented the building of a more "complex" tree structure.).

Our future work we will further validate our method using more data coming from more than one machines, as well as try to refine the feature extraction process trying finer segmentation of the original image.

REFERENCES

[1] W.T. Thomson, M. Fenger, "Current signature analysis to detect induction motor faults" *IEEE Industry Applications Magazine*, July/August 2001, pp. 26-34

[2] J. A. Antonino-Daviu, M. Riera-Guasp, J. R. Folch, and M. Pilar Molina Palomares, "Validation of a new method for the diagnosis of rotor bar failures via wavelet transform in industrial induction machines," *IEEE Trans. Ind. Appl.*, vol. 42, pp. 990-996, 2006.

[3] G. Georgoulas, I.P. Tsoumas, J.A. Antonino-Daviu, V. Climente-Alarcon, C.D. Stylios, E.D. Mitronikas, A.N. Safacas, "Automatic Pattern Identification Based on the Complex Empirical Mode Decomposition of the Startup Current for the Diagnosis of Rotor Asymmetries in Asynchronous Machines," *IEEE Transactions on Industrial Electronics*, vol.61, no.9, pp.4937,4946, Sept. 2014

[4] M. Riera-Guasp, J. A. Antonino-Daviu, M. Pineda-Sanchez, R. Puche-Panadero, and J. Perez-Cruz, "A General Approach for the Transient Detection of Slip-Dependent Fault Components Based on the Discrete Wavelet Transform," *IEEE Trans. Ind. Electron.*, vol. 55, pp. 4167-4180, 2008.

[5] P. Karvelis, I. P. Tsoumas, G. Georgoulas, C. D. Stylios, J. A. Antonino-Daviu, and V. Climente-Alarcon, "An intelligent icons approach for rotor bar fault detection". in Proc. IECON 2013, pp. 5526-5531, 2013

[6] G. Georgoulas, M. Mustafa, I. Tsoumas, J. Antonino-Daviu, V. Climente-Alarcon, C. Stylios, and G. Nikolakopoulos, Principal Component Analysis of the start-up transient and Hidden Markov Modeling for broken rotor bar fault diagnosis in asynchronous machines. Expert Systems with Applications, vol.40, no. 17, pp. 7024-7033, 2013.

[7] M. Fernández Cabanas, M. García Melero, G. Alonso Orcajo, J.M. Cano Rodríguez, J. Solares Sariego. *Maintenance and diagnosis techniques for rotating electric machinery*. Marcombo-Boixareu Editores and ABB Service S.A, Barcelona, 1999.

[8] J. Yoo, J. Yun and S.B. Lee, "Automated Monitoring of High-Resistance Connections in the Electrical Distribution System of Industrial Facilities," IEEE Transactions on Industry Applications, vol. 45, no. 2, Apr. 2009.

[9] N.Y. Utami, Y. Tamsir, A. Pharmatrisanti, H. Gumilang, B. Cahyono and R. Siregar, "Evaluation condition of transformer based on infrared thermography results," *IEEE 9th International Conference on the Properties and Applications of Dielectric Materials*, Harbin, China, July 2009.

[10] B. Li, X. Zhu, S. Zhao and W. Niu, "HV Power Equipment Diagnosis Based on Infrared Imaging Analyzing," *International Conference on Power System Technology,* 2006.

[11] Y. Chou and L. Yao, "Automatic Diagnosis System of Electrical Equipment using Infrared Thermography," International Conference of Soft Computing and Pattern Recognition, 2009.

[12] Y. Han and Y.H. Song, "Condition Monitoring Techniques for Electrical Equipment," IEEE Transactions on Power Delivery, vol. 18, no. 1, January 2003.

[13] B. Bortnem, F. Pray, and P. Grover; "Measuring Shaft Misalignment with Infrared Thermography," P/PM Technology, pp 20-21, February 1995.

[14] S. U. Haq and T. Bashir, "Evaluation of Induction Motor Groundwall Insulation using Infrared Thermography," 2nd International Conference on Emerging Technologies Peshawar, Pakistan, 13-14 November 2006.

[15] M.J. Picazo-Ródenas, R. Royo, J. Antonino-Daviu, J. Roger-Folch, "Use of the infrared data for heating curve computation in induction motors: Application to fault diagnosis," *Engineering Failure Analysis*, Volume 35, 15 December 2013, Pages 178-192

[16] M. Eftekhari, M. Moallem, S. Sadri, Min-Fu Hsieh, "A novel indicator of stator winding inter-turn fault in induction motor using infrared thermal imaging" *Infrared Physics & Technology, Vol. 61, Nov. 2013*, pp. *330-336*

[17] E. Persoon and K. S. Fu, "Shape discrimination using Fourier descriptors," IEEE Trans. Syst., Man, Cybern.,vol. SMC-7, pp. 170–179, Mar. 1977.

[18] N. Ayache and O. D. Faugeras, "Hyper: A new approach for the recognition and positioning of two-dimensional objects," IEEE Trans. Pattern Anal. Machine Intell.,vol. PAMI-8, pp. 44–54, Jan. 1986.

[19] B. Bhanu, "Shape matching of two-dimensional objects," IEEE Trans. Pattern Anal. Machine Intell.,vol. PAMI-6, pp. 137–155, Mar. 1984.

[20] D. Lowe, Object recognition from local scale-invariant features. In International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157, 1999.

[21] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Int'l J. Computer Vision, vol. 2, no. 60, pp. 91-110, 2004.

[22] J. Beis, and D. Lowe, Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. Proc. Conference on Computer Vision and Pattern Recognition (CVPR'97) (Puerto Rico): 1000-1006, 1997.

[23] P. Soille, Morphological Image Analysis: Principles and Applications, Springer; 2nd edition, 2004.

[24] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison Wesley.

[25] J. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

[26] R.. Duda, P. Hart, D. Stork, Patterm Classification, Wiley, 2000.

[27] N. Japkowicz, and M. Shah, Evaluating learning algorithms: a classification perspective. Cambridge University Press, 2011.

[28] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[29] M.. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, vol. 11, pp. 10-18, 2009.

[30] X.Wu, V. Kumar The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.