Geometrical and statistical feature extraction of images for rotation invariant classification systems based on industrial devices

Rodrigo D. C. Silva, George A. P. Thé, Fátima N. S. de Medeiros
Depto. de Engenharia de Teleinformática
Universidade Federal do Ceará
Fortaleza, Brazil
Email: george.the@ufc.br

***Abstract*—In this work, the problem of recognition of objects using images extracted from a 3D industrial sensor is discussed. We focus in 7 feature extractors based on invariant moments and 2 based on independent component analysis, as well as on 3 classifiers (k-Nearest Neighbor, Support Vector Machine and Artificial Neural Network-Multi-Layer Perceptron). To choose the best feature extractor, their performance was compared in terms of classification accuracy rate and extraction time by the k-nearest neighbors classifier using euclidean distance. For what concerns the feature extraction, descriptors based on sorted-Independent Component Analysis and on Zernike moments performed better, leading to accuracy rates over 90.00 % and requiring relatively low time feature extraction (about half-second), whereas among the different classifiers used in the experiments, the suport vector machine outperformed when the Zernike moments were adopted as feature descriptor.**

***Keywords*-Invariant moments, Independent Component Analysis, Support Vector Machine, Multi-Layer Perceptron**

I. Introduction

Automatic visual recognition of objects and people in a scene is a hot research topic worldwide, with many important applications being found in industry (e.g., counting, inspection and quality control), security (e.g., surveillance systems), urban environment (autonomous navigation, collision and accidents avoidance, traffic monitoring), etc [1].

In the industrial automation field, machine vision provides innovative solutions and helps improving efficiency, productivity and quality management, bringing competitiveness for solution providers [2]. As examples of industrial activities which have benefited from the application of machine vision technology on manufacturing processes, it can be cited [3] and references therein.

In general, it is desirable or required for the system to be able to work regardless of translation, rotation or scale transformations, what means achieving good performance in non-structured scenarios, though it may occur at the expense of data processing costs. Goal is, therefore, finding simple and efficient technical solutions.

Thanks to the recent developments in data acquisition, processing, and process control systems, efficiency of many industrial applications has been improved with the help of automated visual processing and classification systems [4]. In this scenario, the classification between different objects (i.e., the ability of label assignment) is usually a complex problem for machines, because it involves many steps, e.g. image acquisition, preprocessing, feature extraction and classification itself. So, although the offer of shop floor equipments has increased a lot (in number and in performance) in the last decades, the majority of industrial computers still has processing and data storage limitations.

On the other side, high-performance computers, nowadays easily available comercially, allows for solving the complexity of the mathematics involved in those mentioned steps. Bridging this gap, the evolution of industrial data networks and integrated communication solutions made possible image-based supervision and on-line control in industry, provided that costy data processing be forwarded to remote stations.

To approach the problem of feature extraction in 2D object recognition, there are some traditional methods which rely mainly on the calculation of invariant moments (Hu, Zernike, Legendre, etc). One important property of the moments is their invariance under affine transformation. Moments are scalar quantities used to characterize a function and to capture its significant features. From the mathematical point of view, moments are projections of a function onto a polynomial basis [5].

More recently, a blind source separation technique named independent component analysis (ICA) has been used in many fields, from electrical power systems [6] to economic modeling [7], since it offers good feature description ability from a reduced set of descriptors [8]. In essence, it represents a given measurement (an image, a speech signal or whatever) as a linear composition of statistically independent components; one could therefore, use the independent components themselves or the coefficients of the linear composition as features of the input raw data.

In this context, in this work the issue of automatic inspection and supervised image classification is considered in both public image datasets and low-resolution images extracted from an industrial sensor. Primary goal is to evaluate the

978-1-4799-0652-9/13/$31.00 ©2013 IEEE

performance of many feature descriptors based on invariant geometrical moments, comparing with feature extraction from independent component analysis. We also aim at investigating different classifiers, in order to point the best solution for image classification.

The paper is organized as follows: in Section II several approaches for feature extraction are described, whereas in Section III the classifiers are briefly reviewed. In the following, in Section IV the experimental setup and the datasets are described. Finally, results and discussion appear in Section V

II. Feature Extraction

According to the literature, feature extraction is the problem of getting, from raw data, relevant information for classification purposes, thus achieving minimal within-class pattern variability while enhancing discrimination between classes. This is accomplished representing each image by a vector containing a set of features. In this section several feature descriptors based on geometrical moments, as well as that based on independent component analysis will be reviewed.

A. Hu moments

In 1962 Hu [9] introduced the concept of moment giving rise to the use of invariant moments and moment functions in the fields of image analysis and pattern recognition. In that seminal paper, seven nonlinear functions which are translation, scale and rotation invariant were introduced for computing the center of mass of a given image, and that of a certain region (in case of a binary mask). As pointed out in [10], the drawback of such approach is that the kernel function of geometric moments of order $(p+q)$ is not orthogonal, leading the geometric moments to suffer from information redundancy, as well as from noise sensitivity for higher-order moments.

B. Zernike moments

The issue of information redundancy can be overcome by the use of orthogonal moments for image representation. This has been known since the early 80's [11] and the Zernike moments were one of the formest approaches adopted (along with Legendre moments, discussed next). In the discrete form, mn-*th* order moment is written as:

$$Z_{mn} = \frac{n+1}{\lambda_N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} V_{nm}(x,y) I_{xy}, \tag{1}$$

where n is the order of the radial polynomial, λ_N is a normalization factor accounting for the amount of pixels inside the unit circle where the basis function, V_{nm} is evaluated and I_{xy} is the image matrix. For better description, refer to [11].

C. Legendre moments

It consists in a recursive relation of the p-*th* order Legendre polynomial, and takes the following discrete form:

$$L_{pq} = \lambda_{pq} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_p(x_i) P_q(y_j) I_{ij}. \tag{2}$$

It is calculated in the interval $[-1, 1]$ and, therefore, the pixels are scaled to the $-1 < x, y < 1$ region, giving rise to the x_i, y_i normalized coordinates. Again, λ_{pq} is a normalization constant and I_{ij} represents the N x N intensity image matrix. Finally, P_q and P_p are Legendre polynomials. For additional details, refer to [11].

D. Fourier-Mellin moments

This is another class of rotation-invariant orthogonal moments, which in the discrete form is calculated as:

$$O_{pq} = \frac{p+1}{\pi} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} I(x_i, y_k) M^*_{pq}(x_i, y_k) \Delta x_i \Delta y_k \tag{3}$$

where x_i and y_k are normalized coordinates of pixels in the region where polynomials are evaluated, M^*_{pq} is the complex conjugate of the orthogonal polynomials and $I(x_i, y_k)$ is the N-sized intensity image matrix.

E. Tchebichef moments

In contrast to the moments described so far, which rely on the definition of a region-of-interest for the evaluation of the discrete form of a given function, Tchebichef moments as introduced by [12] do not require numerical approximations, since the basis functions set is orthogonal in the discrete domain of the image coordinates. These moments are obtained as:

$$T_{pq} = \frac{1}{\rho(p,N)\rho(q,N)} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} t_p(x) t_q(y) I_{xy}, \tag{4}$$

where p and q define the order of the polynomial, ρ is a normalization function dependent on the moment order and on image size, whereas t_p and t_q are Tchebichef polynomials. For further details, refer to [12].

F. Bessel-Fourier moments

It is a more recent class of polar-coordinates based orthogonal moments, introduced in [13]. These are written as:

$$B_{nm} = \frac{1}{2\pi a_n} \sum_k b_{nk} C_{pq}. \tag{5}$$

In this expression, a_n is a normalization constant, b_{nk} is a term accounting for the zeroes of the first-order Bessel function and C_{pq} contains the complex moments dependent on the image matrix of interest. Additional details can be found in [13].

G. Gaussian-Hermite moments

Altough it is not new as a class of continuous orthogonal moments, recently attention has been given in the context of rotation and translation invariance [14]. They are defined as $M = HIH^T$, where I is the image matrix, superscript ^{T}denotes tranpose matrix and H represents the Gaussian-Hermite polynomials having σ scale factor:

$$H(x;\sigma) = \frac{e^{\frac{-x^2}{2\sigma^2}} H_p(\frac{x}{\sigma})}{\sqrt{2^p p! \sqrt{\pi}\sigma}}. \tag{6}$$

In the above equation, H_p is the Hermite polynomial.

H. Independent component analysis

According to [15], it is a statistical signal processing technique whose goal is to linearly decompose a random vector into components which are as independent as possible. The basic definition of ICA considers a set of observations of random variables $x_1(t), x_2(t), ..., x_n(t)$ and the assumption that they are generated as a linear mixture of independent components $s_1(t), s_2(t), ..., s_n(t)$, according to

$$x = A(s_1(t), s_2(t), ..., s_n(t))^T = As, \quad (7)$$

where A is the unknown mixture matrix. For a better explanation on ICA model and underlying requirements, refer to [8]. ICA applications on pattern recognition of rotated images require as training step the random variables to be the training images. Letting x_i to be a vectorized image, we can construct a training image set $x_1, x_2, ..., x_n$, with n random variables which are assumed to be the linear combination of the m unknown independent components s, in such a way that the coefficients are given by the elements of the mixture matrix. When ICA is applied to feature extraction, the columns of A_{train} contain the main feature vectors of the training images, being used therefore as input to the classifier along with the mixture matrix of the image under test, A_{test}.

According to the literature, the efficiency of the ICA algorithm is very dependent on preprocessing steps [8]. Indeed, in applications where rotation invariance is a requirement, we have observed and recently proposed to perform an ordering transformation of the input vectorized images to achieve better ICA feature extraction. This is work is under review, but the technique will be applied to the analysis of the present manuscript for completeness and for comparison to the other feature descriptors. From now on, it will be referred to as *ICAsort*.

III. Classification

After completing the feature extraction, the final stage of any image processing system contains the classification step, in which each sample is labeled or assigned to a new or existent class; in this step, the better the data representation provided by the feature descriptor, the better the assignment step will be. However, also the influence of the classifier itself to the classification efficiency plays a role (efficiency here understood as a measure of its ability to distinguish the interclass similarity whereas bypassing eventual intraclass differences). To accomplish with that, we chose 3 different classification approaches, which will be described next.

A. k-Nearest neighbors classifier

The k-Nearest neighbor is a classifier where the learning is based in analogy. The training samples are formed by n-dimensional vectors, and each element of this group is a point in n-dimensional space.

To determine the class of an element which does not belong to the training set, the k-NN classifier searches for k elements of the training set that are closest to this unknown element, i.e. those whose separation correspond to the smallest distances. These k elements are called k-nearest neighbors. In this article, Euclidean distance is used as metric for evaluating the adjacency [16].

B. Artificial Neural Network

Artificial Neural Network (ANN) can be seen as a parallel distribution process, inspired on how biological neurons process information. It is composed of a large number of highly interconnected processing elements (neurons) working to solve specific problems.

For problems that are not linearly separable, it is possible to efficiently train networks built with intermediate layers, the so-called Multilayer Perceptron network (MLP). A typical MLP network has three main features: the neurons of the intermediate layer have a sigmoid-like activation function, the network has one or more intermediate layers and the network has a high degree of connectivity. [17]

C. Suport Vector Machine

Support Vector Machine (SVM) is a technique for classification and regression that uses a nonlinear mapping to transform the original training data into a higher dimension where a separation hyperplane is better built.

The main idea consists in getting a hyperplane optimum, i.e. hyperplanes which maximize the margin separating the classes, in order to separate training patterns of different classes by minimizing the number of errors in the training group. However, usually the application data is not linearly separable. Thus, the SVM algorithm transforms the nonlinear input characteristics to a space in which linear methods can be applied, thus transforming the data to a space where they can be linearly separable.

Although the training time of even the fastest SVMs can be extremely slow, they are highly accurate, owing to their ability to model complex nonlinear decision boundaries. They are much less prone to over fitting than other methods. [16], [18]

IV. Experimental Setup

A. Public database

As mentioned in the Introduction, the comparison among several feature descriptors and classifiers will be made for different datasets; two of them are public database.

The first database is named dataset A and has 77 images obtained from the database of the Ming Hsieh Department of Electrical Engineering of the University of Southern California. Each image was rotated with 5° step, from 0° to 360°, thus forming 73 samples for every image. Those corresponding to 0° have been used for training, and the remaining, for testing. Some samples are shown in Figure 1.

In the second one, named dataset B, we considered a texture database with different patterns. The Brodatz album available in [19] has 112 texture images, which have been resized from 640 x 640 to 128 x 128 pixels. Here again we rotated images with 5° step, from 0° to 360°, thus forming 73 samples for every image. Once more, those corresponding to 0° have been

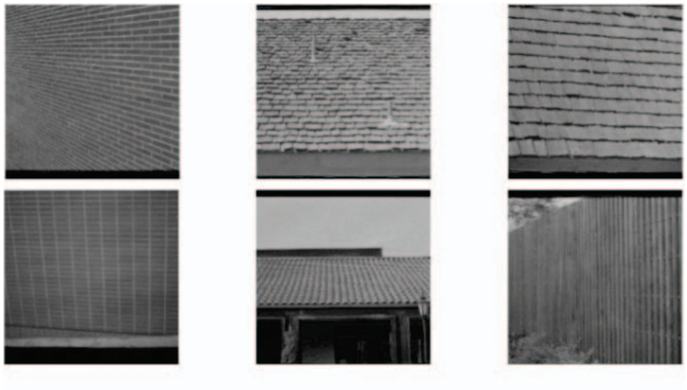

Fig. 1: Samples from University of South California public database.

used for training, and the remaining, for testing. Some samples of this dataset is shown in Figure 2.

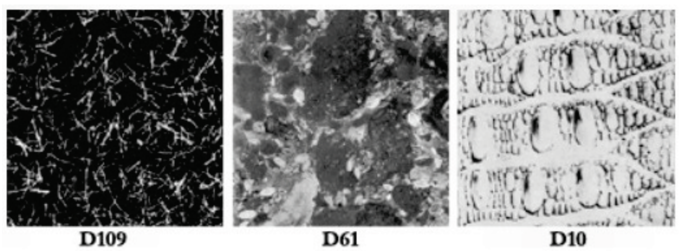

Fig. 2: Samples of Brodatz album.

B. Private database

The third database, dataset C, contains pictures of three small packages, just different in size, which were acquired after randomly rotating the packages on the conveyor belt, until getting 150 samples. This was done in a bad illuminated scenario, as it can be seen in the poor quality of images in Figure 3. The home-made experimental setup for acquisition of dataset C is illustrated in Figure 4.

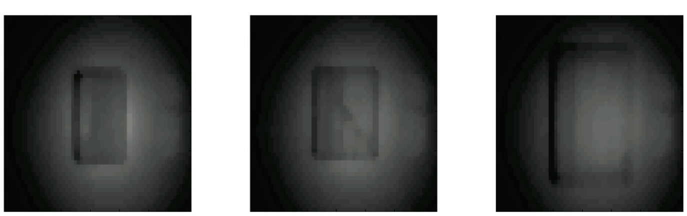

Fig. 3: Pictures of three packages with dimensions 15 x 10.5 x 7.2 cm, 15 x 14 x 6 cm and 21.5 x 16.2 x 9.6 cm, respectively, as acquired from the ifm sensor.

The conveyor is driven by an AC motor with PowerFlex 40P frequency inverter from Allen-Bradley©. This drive is connected to the outputs of the PLC Micrologix 1200©, for remote configuration.

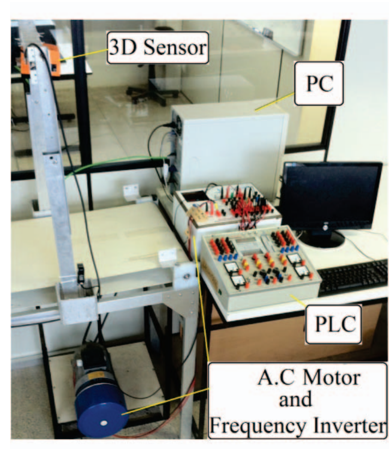

Fig. 4: Experimental apparatus for image acquisition and classification system.

There is also an optical sensor connected to the PLC responsible for triggering the image acquistion whenever an object of interest enters the sensing zone of the industrial camera, 3D effector pmd E3D200 from ifm Electronics©. This is a 50 x 64 resolution camera which uses the time-of-flight (TOF) principle of a matrix of light-emitting diodes to estimate 3D surfaces. It contains Ethernet interface and through remote procedure calls it allows for implementation of real-time applications of classification algorithms remotely.

Communication and data exchange between the running application and the shop floor equipments occurs through Object Linking and Embedding for Process Control Server (OPC) technology. In this solution we use RSLinx Rockwell Automation©to communicate the data managed by the PLC Micrologix 1200©.

V. Results

A. Performance of feature descriptors

To make a fair and clear comparison among the descriptors, initially the feature vectors were presented only to the k-NN classifier using euclidean distance. The classifier was trained and tested 50 times with the same database. The size of the training sets was changed from as low as 10% of the whole available database up to 80%. The experimental results for the dataset C can be seen in the Figure 5 below, whereas those for dataset A and B are plotted in Figures 7 and 6 . In these figures, the following legend was adopted: H - Hu Moments, Z - Zernike, L Legendre, FM - Fourier-Mellin, T - Tchebichef, BF Bessel-Fourier, GH Gaussian-Hermite, ICA Independent Component Analysis and *ICAsort* - ordered Independent Component Analysis (see Section II-H).

Also the time spent for extracting features were evaluated, as well as that for training and running the classifier, and are summarized in Tables I. These results refer to the partition case in which 10% of the dataset was separated for training.

From the table and figures, a number of comments can be made:

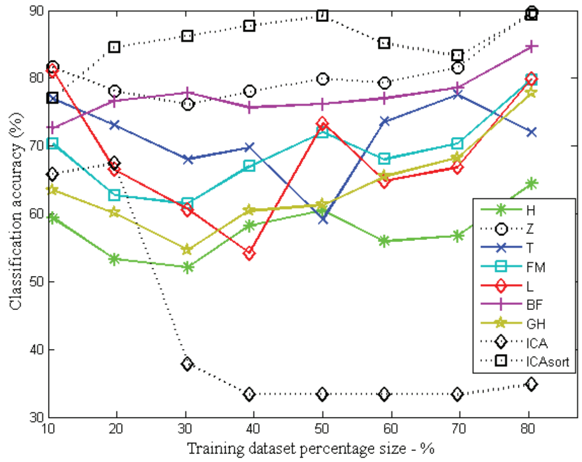

Fig. 5: Average classification accuracy for different feature descriptors, using k-NN and euclidean distance.

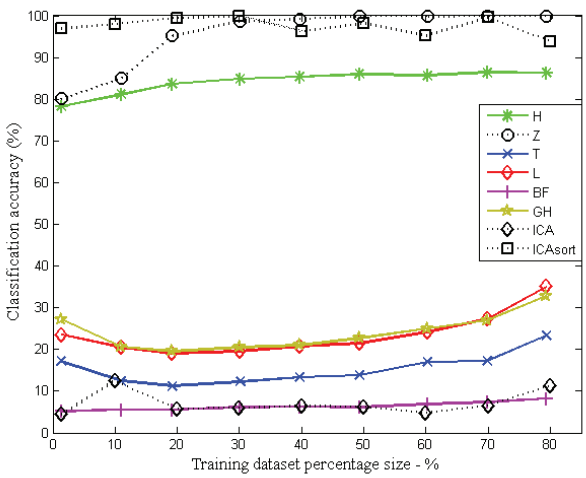

Fig. 6: Average classification accuracy for different feature descriptors, using k-NN and euclidean distance. Obs: Fourier-Mellin omitted for time saving.

a) Zernike moments perform well because they preserve almost all the image information in a few coefficients. The orthogonality of other methods like, for example, the Legendre polynomials has a negative effect when the image is discretized, leading to numerical errors in the calculated moments.

b) Bessel-Fourier has a comparable performance only for the dataset C, which is low-resolution and small database. Furthermore, it seems to be little influenced by the size of the training dataset.

b) Tchebichef moment as feature descriptor leads to irregular classification accuracy, with good performance being achieved in small-sized and large-sized traning datasets. However, for datasets A and B it performs poorly.

c) For what concerns the computational efforts issues, we highlight the fact that Fourier-Mellin moments demand more

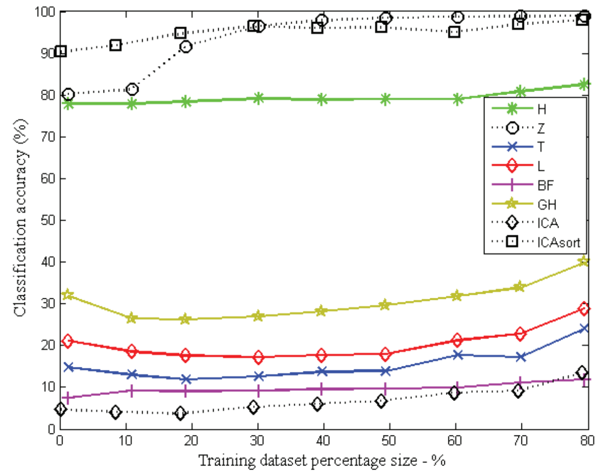

Fig. 7: Average classification accuracy for different feature descriptors, using k-NN and euclidean distance. Obs: Fourier-Mellin omitted for time saving.

TABLE I: Average elapsed times, in seconds, for the various processing steps relative do dataset C.

	Dataset C	
Descriptor	Feature Extraction	Training and classify
H	0.7486	0.0013
Z	0.5782	0.0014
L	0.4580	0.0023
FM	18.49	0.0030
T	0.4485	0.0019
BF	0.5114	0.0009
GH	0.4493	0.0008
ICA	0.6558	0.0011
ICAsort	0.4811	0.0009

resources than others. Its huge running time therefore limits the use in any kind of on-line industrial process. This can be assigned to the large number of summations in the equations, usually solved with iterative loops at the computational level.

d) ICA estimation achieved bad performance in all the experiments carried out. We associate this to the fact it is inherently affected by rotation of the images.

e) Finally, an alternative to this limitation is the *ICAsort* approach, which presented top performance when used as feature descriptor of large datasets even for training sets of reduced size.

Partial conclusion is that feature descriptors that are invariant to rotations in the image plane can be easily constructed using Zernike moments (performed good in most scenarios), but *ICAsort* is a promising alternative.

From the results presented so far, the feature extraction based on Zernike moments as well as on ICA and ICAsort have been chosen for evaluating the performance of classifiers, shown next.

B. Performance of classifiers

In this study of different approaches for classification, we adopted the following scenario:

a) SVM is implemented with polinomial kernel d=1

b) ANN-MLP has sigmoid (tanh) as activation function, and the number of neurons in the hidden and output layers equal 7 and 5, respectively; convergence is ensured by the backpropagation training algorithm.

In Figure 8 we plot the mean accuracy for different combinations of descriptor/classifier as the training dataset size is increased. In this study, only images from dataset C were used. Some comments can be drawn from that figure:

a) using ICA as feature extractor makes the classification very dependent on the training dataset size; a significant sensibility is observed for all the classification approaches.

b) the same comment applies for the *ICAsort* descriptor when SVM or neural networks are used as classifiers. The only good exception is the k-NN approach, which shows nearly regular accuracy in the range of training dataset size studied.

c) Also the feature descriptor based on the Zernike moments seems to be less sensitive to the training set when k-NN is used. For what concerns the SVM and neural network based classifiers, however, there is a positive trend along the training dataset size, and the approach based on Zernike moments + SVM classifier revealed to be superior in the present analysis.

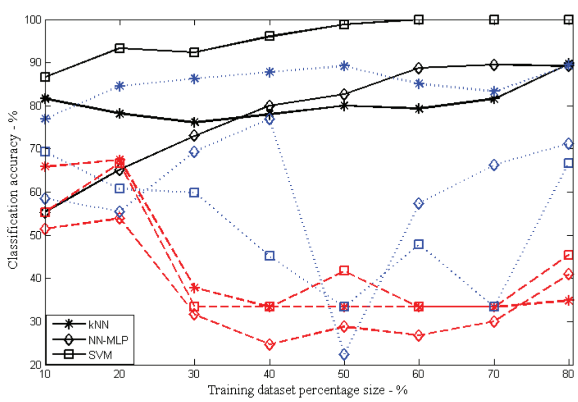

Fig. 8: Performance of classifiers for different alternatives of feature extraction.

VI. Conclusion

This paper reported a comparative study between seven invariant moments (Hu, Zernike, Legendre, Fourier-Mellin, Tchebichef, Bessel-Fourier and Gaussian-Hermite) and Independent Component Analysis as feature descriptors of images from differents databases. Furthermore, a short comparison between three approaches for classification was made, namely the k-NN classifier, a Neural-Network based classifier and a Support Vector Machine.

The study of the feature extraction step revealed that Zernike moments and *ICAsort* are good candidates for feature description, with a slight advantage of *ICAsort* when k-NN is adopted as classifier.

The study of the classifier, in turn, revealed the superiority of the SVM when the Zernike moments are used as feature descriptor.

To sum up, according to the present study we may state that whenever the feature extraction comes from the *ICAsort* algorithm, the k-NN should be preferred as classifier. Should the feature extraction be performed with Zernike moments, so a Support Vector Machine classifier is recommended instead.

Acknowledgement

Authors acknowledge CAPES and FUNCAP (PP1-0033-00032.01.00/10) for financial support. Authors also thank Rockwell Automation do Brasil, for the support through the Scientific and Technical Cooperation Agreement with the Federal University of Ceará. Authors finally thank NUTEC, for administrative facilities.

References

[1] E. B. Corrochano, *Handbook of Geometric Computing Applications in Pattern Recognition, Computer Vision, Neural Computing and Robotics*. Heidelberg: Springer, 2010.

[2] E. N. Malamas, E. G. M. Petrakis, M. Zervakis, L. Petit, and J. D. Legat, "A survey on industrial vision systems, applications and tools," *Image and Vision Computing*, vol. 21, 2003.

[3] D. Sankowski and J. Nowakowski, *Computer Vision in Robotics and Industrial Applications*. Singapore: World Scientific, 2014, vol. 3.

[4] M. A. Selver, O. Akay, F. Alim, S. Bardakci, and M. Olmez, "An automated industrial conveyor belt system using image processing and hierarchical clustering for classifying marble slabs," *Robotics and Computer-Integrated Manufacturing*, vol. 27, 2011.

[5] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. Chichester: John Wiley & Sons.

[6] M. A. A. Lima, A. S. Cerqueira, D. V. Coury, and C. A. Duque, "A novel method for power quality multiple disturbance decomposition based on independent component analysis," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 593–604, 2012.

[7] T.-Y. Lin and S.-H. Chiu, "Using independent component analysis and network dea to improve bank performance evaluation," *Economic Modelling*, vol. 32, pp. 608–616, 2013.

[8] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Components Analysis*. Canada: John Wiley & Sons, Inc., 2001.

[9] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on information theory*, 1962.

[10] D. Sridhar and I. V. M. Krishna, "Face recognition using tchebichef moments," *International Journal of Information & Network Security*, vol. 4, pp. 243–254, 2012.

[11] M. R. Teague, "Image analysis via the general theory of moments," *Journal Optical Society American*, vol. 70, no. 8, pp. 920–930, 1980.

[12] R. Mukundan, S. H. Ong, and P. A. Lee, "Image analysis by tchebichef moments," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1357–1364, 2001.

[13] B. Xiao, J. F. Ma, and X. Wang, "Image analysis by bessel-fourier moments," *Pattern Recognition*, vol. 43, 2010.

[14] B. Yang and M. Dai, "Image analysis by gaussian-hermite moments," *Signal Processing*, vol. 91, 2011.

[15] L. Fan, F. Long, D. Zhang, X. Guo, and X. Wu, "Applications of independent component analysis to image feature extraction," in *Proceedings of the Second International Conference on Image and Graphics*, vol. 4875, 2002, pp. 471–476.

[16] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Waltham: Elsevier, 2006.

[17] A. H. Kulkarni and S. A. Urabinahatti, "Performance comparison of three different classifiers for hci using hand gestures," *International Journal of Modern Engineering Research*, vol. 2, no. 4, pp. 2857–2861, 2012.

[18] R. E.-Maleki, A. Rezaei, and B. M.-Bidgoli, "Comparison of classification methods based on the type of attributes and sample size," *JCIT*, vol. 4, no. 3, pp. 94–102, 2009.

[19] A. Safia and D. He, "New brodatz-based image databases for grayscale color and multiband texture analysis," *ISRN Machine Vision*, no. id 876386, 2013.