Automatic recognition of industrial tools using artificial intelligence approach

Tomasz Les [a], Michal Kruk [c], Stanislaw Osowski [a,b,*]

[a] Warsaw University of Technology, Warsaw, Koszykowa 75, Poland
[b] Military University of Technology, Warsaw, Kaliskiego 2, Poland
[c] University of Life Sciences, Warsaw, Nowoursynowska 165, Poland

ARTICLE INFO

ABSTRACT

The paper presents an automatic approach to the recognition of the industrial tools on the basis of their image registered by the camera. The solved tasks include: the automatic localization of the tools in the image, preprocessing of the image (binarization, noise removal, filling the holes, normalization, etc.), generation of the numerical descriptors, verification of descriptors in the role of the diagnostic features, selection of the features and final stage of classification of the tools. The efficiency of the developed system has been verified on the example of exemplary set of industrial tools and the results of this verification are presented and discussed in the paper.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Machine vision systems are being increasingly used for sophisticated applications such as recognition and classification of different processes. An important milestone in the development of "intelligent" inspection systems has been the rapid growth of computing power in recent years, coupled with the idea that we could successfully emulate the low-level mechanisms of the brain. Thanks to the methodology offered by the artificial intelligence methods, they are able to solve difficult computational problems arising in the task of object recognition.

The natural human visual system has the ability to recognize an object despite changes in the object's position in the input field, its size, or its orientation. For many industrial applications involving classification of elements under recognition, the machine vision systems must also have this ability. Many industrial applications of machine vision allow objects to be identified and classified by their boundary contour or silhouette (Kim & Nam, 1995; Nabhani & Shaw, 2002). An example of it is a robot recognition of industrial parts and tools in the factor environment. In most assembly or sorting lines several different types of tools of the shapes known in advance are handled. Their recognition belongs to the problem of object recognition or pattern matching (Duda, Hart, & Stork, 2003).

This work presents and investigates the application of machine vision methods in the processing of single camera multi-positional images for 2D object recognition and classification. The boundary contour information was chosen as the basis for representing the chosen industrial tools at their recognition. The appropriately preprocessed shape of the object may form the set of diagnostic features, used as the input information to the classifier, responsible for the final recognition of the object.

In our approach to the recognition/classification of industrial components we divide the process into two major stages. The first one is the extraction of the component of interest from the registered image and characterization of it by the numerical features well representing its shape. In the second stage the numerical descriptors, called also the diagnostic features, are put to the input of automatic classifier, responsible for the final recognition. Many recent solutions of classifiers, including artificial neural networks may perform the role of final recognition. In this work we will apply two most efficient solutions: the random forest of decision trees and Support Vector Machine (SVM).

The numerical experiments will be done on the set of simple industrial tools differing by the shape, size, location and orientation in the input field. The preprocessing steps will extract the component from the image, convert them to the normalized features independent on their location, orientation and size. After the selection process we get the most important features, which are used as the input information to the classifier, performing the final recognition.

2. The preprocessing stage

The image of the input field may contain the interesting object represented in general by RGB color components and some smaller elements representing the noise. The main task of this stage is to localize the object and preprocess it in the way enabling to get its numerical characterization.

---

* Corresponding author at. Warsaw University of Technology, Warsaw, Koszyko-wa 75, Poland. Tel.: +48 222347235.
E-mail address: sto@iem.pw.edu.pl (S. Osowski).

2.1. Binarization of the image

The binarization of the image is the process of converting the image into the binary one, in which each pixel has one of two possible intensity levels: either one or zero. The first step is to convert the color into gray levels. This is done for each pixel in the position (x,y) by using standard relation (Gonzalez & Woods, 2011)

$$f(x,y) = 0.3R(x,y) + 0.59G(x,y) + 0.11B(x,y) \qquad (1)$$

where R, G and B represent the three color characterization of the pixel. In the second stage the 256 gray levels are converted to a black and white image by choosing the appropriate threshold value, and classifying all pixels with values above this threshold as white, and all other pixels as black. The important problem is to select the correct threshold. It depends on the lighting conditions at the image registration as well as on the type of the image. The threshold value may be selected on the way of many trials or by applying the adaptive method, for example the Otsu method (Otsu,1979).

2.2. Localization of an object

The next step of processing is the localization of the tool in the input field image. The basic assumption is that the tool occupies the largest area in the field. Therefore we scan the whole field searching for the largest compact region. This algorithm is called the grain growth (Gonzalez & Woods, 2011; Soille, 2003) and is relied on finding all pixels forming compact region.

The algorithm starts form the middle of the image looking for the black pixel. Next we move to all possible black pixels adjacent to the already found, increasing in this way the grain. The largest found compact region composed of only black pixels is taken as our object of interest. All other pixels of the image are treated as the background and changed to the white color. The last step is to fill the holes in the discovered object. This is done by applying the morphological operation of opening and closing (Soille, 2003).

Fig. 1 illustrates the process of grain growth applied for discovering the shape of a grinder. The first image (a) illustrates 1% advancement, the second (b) 40% of advancement of grain growth, the third (c) 100% of grain growth and the last one (d) the shape of the tool after filling the holes.

2.3. Definition of numerical descriptors of the image

After recognizing the shape of the object the next step is to define the numerical descriptors characterizing the object. In our solution we have relied on the geometrical descriptors. The basis of this description is formed by the geometrical characterization of the object, especially the real and convex areas of the object, the real and convex perimeters, as well as box counting dimension (Schroeder, 2006).

The real area of the object is treated as the total number of black pixels forming the object. The perimeter is the number of boundary pixels of the object. The convex parameters (area and perimeter) have been defined using the concept of convex envelope fixed on the boundary points of the shape of the object (Cormen, Leiserson, Rivest, & Clifford, 2009).

The box counting dimension belongs to one of the measures frequently used in fractal theory. It is a method of analyzing data of complex patterns by breaking the object into smaller and smaller pieces called boxes, and analyzing the pieces at each smaller scale. The aim of this process is to examine how observations of detail change with scale. In principle it measures how the length of the complex curve is changing when the measurement is performed with the increased accuracy (Schroeder, 2006).

Let us assume that the object under characterization is placed on the surface covered with the set of regular cubes (squares in 2-dimensional space) of the size ε. Then we count the number of cubes that contain any fragment of the object. This number is evidently dependent on the size of ε. Let us denote it as $N(\varepsilon)$. Changing the size ε we get different values of $N(\varepsilon)$, generally increasing when ε is decreased.

In practice instead of ε we use the number s^2 of elementary cubes, each of the size ε, fixed on the considered analyzed squared area. The value of s is inversely proportional to ε. In this way the definition of the box counting dimension d can be written in the form

$$d = \lim_{s \to \infty} \frac{\log(N(s))}{\log(s)} \qquad (2)$$

The value d, interpreted as the slope of the curve $\log(N(s))$ with respect to $\log(s)$, represents the box-counting dimension, characterizing the complexity of the curve. To use the box-counting notion we have to define only one parameter – the number of boxes, dependent on s. This number defines also the size of the elementary binary matrix $s \times s$. We check if any fragment of the curve enters each box. If yes, we put one in proper location of the matrix, in other case zero. In this way we get the number $N(s)$. Using the relation (2) we get the value of the box counting dimension. The box counting dimension can form the direct numerical descriptor, since according to the definition it represents the relative value.

On the other hand from the analysis of the images it is evident that the basic geometrical parameters (area and perimeter) could not be used as the direct descriptors, since their values change with the size of the object. Instead we have use some derivatives of them, defined in a relative way. Let us denote the directly measured parameters by the following symbols:

A – real area of the object
A_c – convex area of the object
P – perimeter of the object
P_c – convex perimeter of the object
B – box counting dimension

On the basis of them we have defined the following descriptors, being the potential diagnostic features (Kruk, Osowski, & Koktysz, 2007)

- Circularity ratio

$$F_c = \frac{4\pi A}{P} \qquad (3)$$

Fig. 1. The illustration of the process of grain growth for discovering the grinder in the input field: (a) 1% of advancement, (b) 40% of advancement, (c) 100% of advancement, (d) the final shape after filling the holes.

- Convex circularity ratio

$$F_{cc} = \frac{4\pi A_C}{P_C} \tag{4}$$

- Compactness factor

$$F_h = \frac{P^2}{A} \tag{5}$$

- Convex compactness factor

$$F_{ch} = \frac{P_c^2}{A_c} \tag{6}$$

- Corrugation factor

$$F_{ce} = \frac{P_c}{P} \tag{7}$$

- Shredding factor

$$F_r = \frac{A}{A_c} \tag{8}$$

- The ratio of B and P

$$F_{bp} = \frac{B}{AP} \tag{9}$$

- The ratio of B and P_c

$$F_{bpc} = \frac{B}{P_c} \tag{10}$$

- The ratio of B and A

$$F_{ba} = \frac{B}{A} \tag{11}$$

- The ratio of B and A_c

$$F_{bac} = \frac{B}{A_c} \tag{12}$$

These 10 descriptors and the box counting dimension form the set of 11 potential diagnostic features used by the classifier to recognize the objects.

2.4. Selection of the features

The automatically generated descriptors may have different impact on the classification process (Duda et al., 2003; Guyon & Elisseeff, 2003). Some of them may occupy the same values for different classes and some represent the noise from the point of view of pattern recognition. Good feature should be characterized by the stable values for samples belonging to the same class and at the same time they should differ significantly for different classes. Thus the important problem in the classification and machine learning is to find out the features of the highest importance for the problem solution. Observe that the elimination of some features leads to the reduction of the dimensionality of the feature space and improvement of performance of the classifier in the testing mode for the data not taking part in learning.

From many known techniques of feature selection (Duda et al., 2003; Guyon & Elisseeff, 2003) like principal component analysis, analysis of correlation existing among features, correlation between the features and the classes, application of feature ranking by applying the linear SVM, the analysis of mean and variance of the features belonging to different classes, we have chosen the last one.

The variance of the features corresponding to members of one class should be as small as possible. On the other hand, to distinguish between different classes, the positions of means of feature values for the data belonging to different classes should be separated from each other as much as possible. Both requirements are combined together to form the discrimination coefficient $S_{AB}(f)$ defined for the feature f at recognition of two objects belonging to different classes A and B

$$S_{AB}(f) = \frac{|c_A(f) - c_B(f)|}{\sigma_A(f) + \sigma_B(f)} \tag{13}$$

In this definition c_A and c_B are the mean values of the feature f in the class A and B, respectively. The variables σ_A and σ_B represent the standard deviations determined for the respective class. The large value of $S_{AB}(f)$ indicates good potential separation ability of the feature f for these two classes. On the other side its small value means that this particular feature is not good for the recognition between classes A and B. The set of descriptors of highest values of discrimination coefficients form the optimal set of features.

3. Classification systems

The selected set of features is the diagnostic information put to the input of the classifiers. To get the best results of pattern recognition we have to apply the most efficient classifiers. According to the actual knowledge to the best solutions of the classifiers belong the Support Vector Machine (Schölkopf & Smola, 2002; Vapnik, 1998) and Breiman random forest of the decision trees (Breiman, 2001). These two systems of classifiers implemented in Matlab (Matlab user manual with toolboxes, 2012) will be applied and studied in the paper.

Basically the SVM is a linear machine, working in the high dimensional feature space formed by the non-linear mapping of the N-dimensional input vector **x** into a L-dimensional feature space ($L > N$) through the use of a kernel function $K(\mathbf{x}, \mathbf{x}_i)$. It is known as an excellent classifier of good generalization ability (Vapnik, 1998). The learning problem of SVM is formulated as the task of separating the learning vectors into two classes of the destination values either $d_i = 1$ (one class) or $d_i = -1$ (the opposite class), with the maximal separation margin. The separation margin formed in the learning stage according to the assumed value of the regularization constant C provides some immunity of this classifier to the noise, inevitably contained in the testing data.

The great advantage of SVM is the unique formulation of learning problem leading to the quadratic programming with linear constraints, which is easy to solve. The SVM of the Gaussian kernel has been used in our application. The hyperparameters σ of the Gaussian function and the regularization constant C are usually adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one at the validation data sets. The typical values of these parameters for the normalized input data are $\gamma \approx 1$ and $C \approx 1000$. To deal with a problem of many classes the one against one or one against all approaches working on a principle of the majority voting (Schölkopf & Smola, 2002) are usually applied. In our solution we have applied the one against one approach, since this approach lead to the better total results of recognition.

The Breiman (Breiman, 2001) random forest is an ensemble of many decision trees and outputs the class pointed by the majority of the individual trees. The method combines "bagging" idea and the random selection of features for each node in order to construct a collection of decision trees with controlled variation. Each tree in the forest is constructed in a way providing the highest degree of independence.

Assume that the number of training cases is p, and the number of input variables in the classifier be N. Denote by m the number of input variables to be used to determine the decision at a node of the tree; m should be less than N. Choose a training set for the tree by choosing n times (with replacement) from all p available training cases. Use the rest of the cases to estimate the error of the tree, by predicting their classes. For each node of the tree, choose randomly m variables on which to base the decision at that node. Determine the best split based on these m variables in the training set. Allow each tree to be fully grown and not pruned.

The class prediction of a new sample is done by pushing it down the set of trees. Each tree assigns the label of the training sample in the terminal leave it ends up in. The same procedure is iterated over all trees in the ensemble, and the majority of votes of all trees in the forest dictates the class.

4. The results of numerical experiments

4.1. Data base

In the numerical experiments we have recognized 9 classes of the industrial tools. The recognized classes include: class 1 – setsquares, class 2 – magnifying glasses, class 3 – hammers, class 4 – boxes, class 5 – screwdrivers, class 6 – pliers, class 7 – drills, class 8 – discs for grinders, class 9 – grinders. The single exemplary representatives of each class are presented in Fig. 2.

In the first set of experiments of the pattern recognition we have represented each class by 33 individuals, differing significantly from each other. Fig. 3 presents the exemplary set of tools representing the class of hammers. As it is seen they differ significantly by the shape, size, proportion of parameters, orientation, position in the viewing field, etc.

In the second set of experiments we have checked how our automatic system is resistive to different types of noise disturbing the individual image. This time we have assumed that the system recognizes one chosen representative of each tool, represented in a different scale, various positions in the viewing field, at the presence of some cracks, traces of dust of different shape, existence of blurs, etc. The number of representatives of each class was also equal 33. The exemplary set of grinder's discs used in experiments are presented in Fig. 4.

4.2. The results of the first set of experiments

The aim of the first experiments was to check how the developed system is resistive to the changes of the shape of the tools, such as that presented in Fig. 3 for hammers. In the first stage of preprocessing, after extraction of the shapes of each tool, the most important was the generation and selection of the diagnostic features defined in Section 2.3. We have checked the discriminative strength of each descriptor for all combinations of classes (36 pairs of 2-class problems). For each pair of classes the value of Fisher discriminant coefficient has been calculated and then averaged over all pairs. The cumulative diagnostic value of the descriptor f_k ($k = 1, 2, \ldots, 11$) is defined as

$$S(f_k) = \frac{1}{36} \sum_{\substack{i,j=1 \\ i \neq j}}^{11} S_{ij}(f_k) \tag{14}$$

Fig. 5 presents the distribution of the values of $S(f_k)$ for the succeeding features, averaged over all combinations of classes. As it is seen the highest cumulative value represents shredding factor, which is over two times better than the next one (the convex compactness factor). The least discriminative values are associated with all four ratios of box counting dimensions with respect to the geometrical
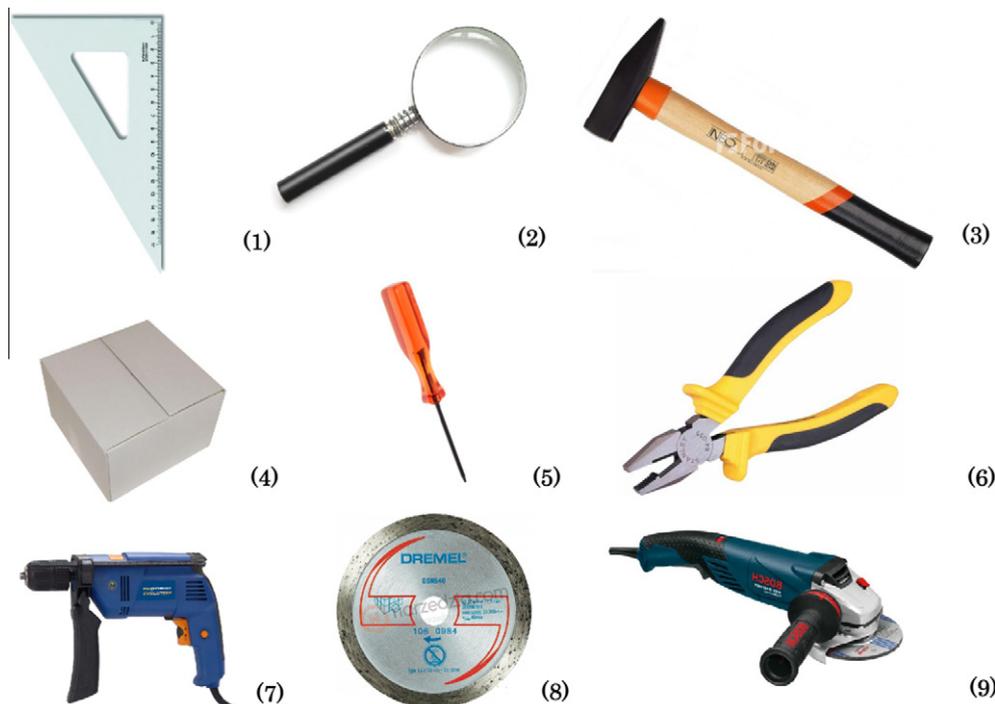

Fig. 2. The examples of tools used for recognition in the numerical experiments (each tool is represented by a single example).

Fig. 3. The set of 33 images of hammer used in the experiments.

parameters. However the box counting dimension B itself is reasonably good.

To check the quality of such values of the discriminative measures of the diagnostic features we have mapped the data into a 2-dimensional system represented by two best descriptors. Fig. 6 presents the distribution of data in two dimensional space formed by F_r and F_{ch} (two best descriptors).

We can see the interesting distribution of the data points. The representatives of classes are placed in different regions of the plane, usually in the compact sets. However, some classes interlace with the other, making their recognition impossible. It is evident, that the recognition of classes using only two diagnostic features is impossible. We have to use richer set of descriptors. To get their optimal quantity we have made introductory experiments of recognition at application of varying number of the features arranged in the decreasing order of their discriminative ability. As the classifier we have applied the Gaussian kernel SVM working in one against one mode. Two third of the data has been used in learning and the rest (1/3 of data) in testing. As a results of these experi-

ments we have found the optimal number of descriptors is equal 5 (F_r, F_{ch}, F_h, F_{ce} and B).

Next we have performed the 10 fold cross validation experiments of recognition using these 5 diagnostic features as the input information for the Gaussian kernel SVM operating in one against one mode. The data was divided into 10 parts. Nine parts were used in learning and the last one in testing. In each experiment the testing part was exchanged with one of the learning parts. The results of testing were averaged. The mean value of the relative testing error was equal 8.42%. Table 1 depicts the confusion matrix of the recognition, presented in the form of recognition results concerning the testing data not taking part in learning in all 10 cross validation experiments.

The same experiments have been repeated using the random forest as the classifier. Two third of data points were used in learning and 1/3 of testing. The best results of recognition have been obtained by assuming m = 3 variables used to determine the decision at a node. The best results of recognition obtained at application of 30 decision trees in the forest was this time worse than that at

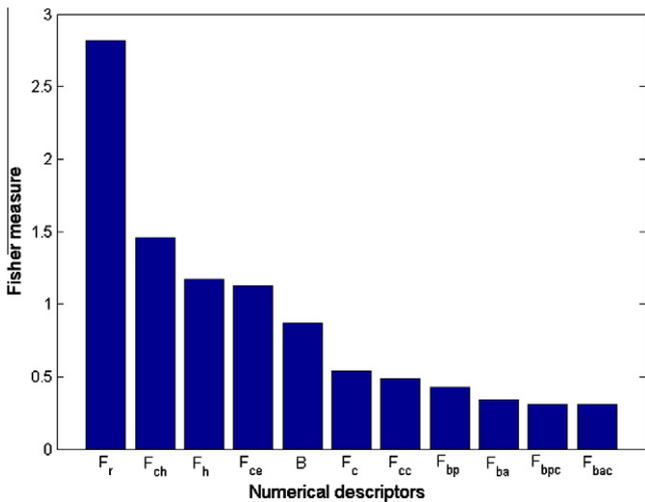**Fig. 4.** The exemplary set of noisy images of discs of grinder used in the second type of experiments.

Fig. 5. The averaged cumulative values of the Fisher discriminant measure for the succeeding features in the first set of experiments.
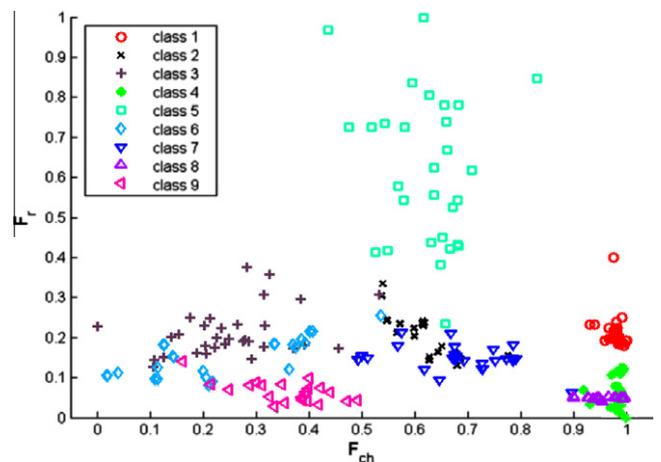

Fig. 6. The distribution of data points in the coordinate system represented by two most important features.

Table 1
The confusion matrix of recognition of 9 classes of industrial tools.

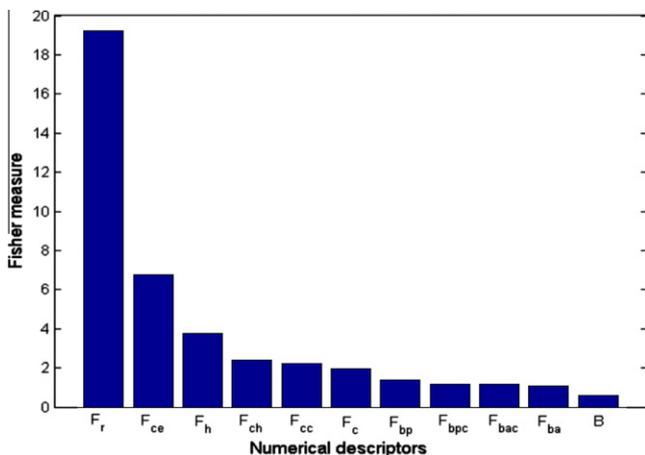| Classes	1	2	3	4	5	6	7	8	9
1 (Setsquares)	33	0	0	0	0	0	0	0	0
2 (Magnifying glasses)	0	29	0	0	0	1	3	0	0
3 (Hammers)	0	2	30	0	0	1	0	0	0
4 (Boxes)	0	0	0	29	0	0	0	4	0
5 (Screwdrivers)	0	1	0	0	31	1	0	0	0
6 (Pliers)	0	1	1	0	0	30	0	0	1
7 (Drills)	1	2	1	0	0	1	28	0	0
8 (Discs for grinders)	0	0	0	3	0	0	0	30	0
9 (Grinders)	0	0	0	0	0	1	0	0	32

application of SVM. The mean value of the relative testing error was equal 10.46%.

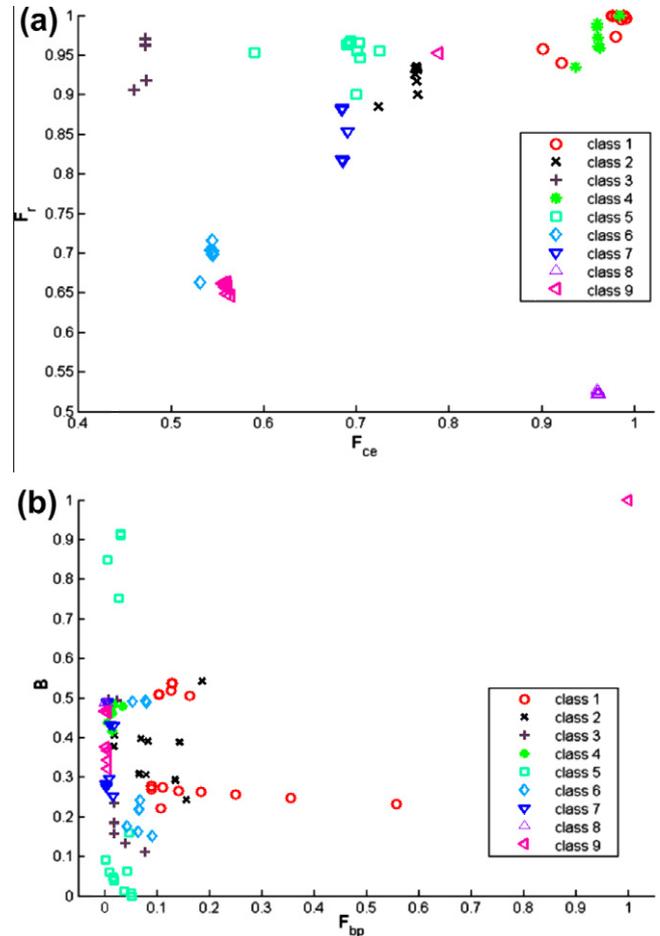### 4.3. The results of the second set of experiments

The second set of experiments aimed on checking the sensitivity of the developed system on the existence of the noise and other disturbances existing in the image of the tools. This time the original shape of the tool was unique and the disturbances have been introduced artificially by us. The exemplary set of noisy images of discs of grinder used in the second type of experiments was shown in Fig. 4. The other tools have been also disturbed in a similar fashion.

Similarly to the previous experiments for each pair of classes we have calculated the value of Fisher discriminant coefficients and then averaged over all pairs. The cumulative diagnostic value of the succeeding descriptors arranged in the decreasing order are presented in Fig. 7. Once again the highest cumulative value represents shredding factor, which is more than three times higher than the next one (the compactness factor). The least discriminative values are associated with the box counting dimension and its ratio to the geometrical parameters. To check the discriminative quality of two best descriptors we have presented the distribution of classes in a 2-dimensional system (Fig. 8(a)). As it is seen most of the representatives of the classes form the compact clusters placed very close to each other. Only single samples of different classes are interlaced with each other (one sample of class 9 placed close t class 2, the representatives of class 1 very close to class 4).

To check how reliable is application of Fisher measure of discrimination ability of descriptors we have compared the data distribution at application of only two worst features (B and F_{ba}). It is shown in Fig. 8(b). As it is seen this time the representatives

Fig. 8. The distribution of data points in the coordinate system represented by (a) two most important features: shredding and corrugation factors and (b) two least significant descriptors.

of all classes are interlaced with each other making the recognition of classes very hard.

The experiments of recognition have been made by applying the Gaussian kernel SVM working in one against one mode and random forest of decision trees at $m = 4$, found to be the best. Te experiments have been conducted in the same way as in the first case (two third of the data used in learning and the rest in testing, experiments arranged in a 10-fold cross validation). The best results have been obtained at application of the first four descriptors: F_r, F_{ce}, F_h, F_{ch} used as the diagnostic features.

The averaged results of testing obtained at application of SVM and random forest have been compared. This time the mean value of the relative testing error was much smaller and equal 1.12% for SVM and only 0.67% for random forest. Table 2 shows the

Table 2
The confusion matrix of recognition of 9 classes of industrial tools in the second set of experiments.

Classes	1	2	3	4	5	6	7	8	9
1 (Setsquares)	33	0	0	0	0	0	0	0	0
2 (Magnifying glasses)	0	33	0	0	0	0	0	0	0
3 (Hammers)	0	0	32	0	0	1	0	0	0
4 (Boxes)	0	0	0	33	0	0	0	0	0
5 (Screwdrivers)	0	0	0	0	33	0	0	0	0
6 (Pliers)	0	0	0	0	0	33	0	0	0
7 (Drills)	0	0	0	0	0	0	33	0	0
8 (Discs for grinders)	0	0	0	0	0	0	0	33	0
9 (Grinders)	0	1	0	0	0	0	0	0	32

Fig. 7. The averaged cumulative values of Fisher discriminant for the succeeding features in the second set of experiments.

confusion matrix of the pattern recognition at using the random forest as the classifier system. They are presented in the form of recognition results of testing data not taking part in learning in all 10 cross validation experiments.

As expected we have got much smaller error of recognition, since the shape of each tool was this time known in advance and the only differences among the samples have been caused by the effect of noise, artificially introduced to the original images.

5. Conclusions

The paper has presented the automatic approach to the recognition of the industrial tools by using the methods based on the artificial intelligence and advanced image processing. The boundary contour information was chosen as an effective method of representing the industrial components under recognition.

The starting point of the recognition procedure is the camera image of the tools. The succeeding stages of developed procedure included: the automatic localization of the tools in the image, pre-processing of the image such as the binarization, noise removal, filling the holes, and normalization, generation of the numerical descriptors characterizing the shape of the tools, verification of descriptors in the role of the diagnostic features, selection of the features and final stage of classification of the tools. Two different solutions of the final classifier system have been tried. One was based on Support Vector Machine and the second applied the random forest of the decision trees. Both belong to the most efficient automatic classification systems.

The efficiency of the developed system has been verified on the example of 9 exemplary sets of the industrial tools. The tools differing significantly with a shape, size and location have been considered first in the recognition process. The numerical results of this verification have been presented and discussed in the paper. The next experimental set assumed recognition of the known in advance shape of the tools under the noisy environment. Experimental results show the good performance of the proposed system both in a noise-free and the noisy environment.

References

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Cormen, T. H., Leiserson, C. E., Rivest, R., & Clifford, C. S. (2009). *Introduction to algorithms*. Boston: MIT Press.

Duda, R. O., Hart, P. E., & Stork, P. (2003). *Pattern classification and scene analysis*. New York: Wiley.

Gonzalez, R. C, & Woods, R. E (2011). *Digital image processing*. Upper Saddle River: Prentice Hall.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

Kim, H., & Nam, K. (1995). Object recognition of one-DOF to tools by back-propagation neural net. *IEEE Transaction on Neural Networks, 6*, 484–487.

Kruk, M., Osowski, S., & Koktysz, R. (2007). Numerical characterization of the images of prostate cancer for recognition of Gleason scale. *Przeglad Elektrotechniczny, 83*, 227–230.

Matlab user manual with toolboxes (2012). Natick: MathWorks.

Nabhani, F., & Shaw, T. (2002). Performance analysis and optimization of shape recognition and classification using ANN. *Robotics and Computer-Integrated Manufacturing, 18*, 177–185.

Otsu, N. (1979). A threshold selection method from grey-level histograms. *IEEE Transaction on Systems, Man, and Cybernatics, 9*, 62–66.

Schölkopf, B., & Smola, A. (2002). *Learning with kernels*. Cambridge MA: MIT Press.

Schroeder, M. (2006). *Fractals, Chaos, power laws*. New York: W.H. Freeman and Company.

Soille, P. (2003). *Morphological image analysis, principles and applications*. Berlin: Springer.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.