Computer Science Journal of Moldova, vol.**25**, no.1(73), 2017

Fuzzy Rule-based System for Pattern Recognition and Automated Classification

Mariana Rusu, Marius-Dan Zbancioc

Abstract

The sorting method contains a fuzzy system that classifies waste into three classes: dangerous (toxic, flammable, harmful to the environment), recyclable and undetermined. To increase the decision weight since some waste is perilous (eg. batteries, spray bottle under pressure, oil drums, etc.), we implement a shape-signature identification algorithm.The extraction of objects contour in order to obtain the signature is done with contour tracking algorithm. The appropriate determination of contour also enables efficient calculation of shape and size features used in fuzzy system. The elaborated system aims to detect dangerous wastes that arrive among the recyclable waste because of people ignorance.

Keywords: Automated sorting, border tracing algorithm, shape-signature identification, fuzzy system.

1 Introduction

Improper waste management seriously affects the environment not only by local pollution but also contributing to global greenhouse effect gas emissions.

It is known mechanical, aeraulic sorting, infrared spectroscopy and X-ray detection for plastic (it's even possible to differentiate types PVC, PET, etc.), but the most effective sorting is currently the manual sorting. The sorting staff can separate bottles of different color, paper of different quality, dangerous waste etc. This method is costly both in cost and in time. One possibility to increase the efficiency is the use of semiautomatic Decision Support Systems [1], [2], [3]. There are known implementations of optical sorting through the application of algorithms for digital images recognition [4], [5], [6]. The main objective of designed SSD for waste sorting depends on the critical issues facing a certain society, country. Many researchers have applied waste management options: material recovery through recycling / reuse (paper, plastic, metals, etc.), incineration, composting and storage of perilous waste.

The automated sorting method proposed in this paper includes four basic steps. At the first step we identify if there is a danger symbol on the label using normalized correlation coefficient (CCN) and a genetic algorithm (GA). If the label contains one of the symbols of danger, the packaging is considered dangerous and is removed from the packaging tape for recycling. The second step consists in extracting the characteristics of packaging and includes a contour tracking algorithm and an algorithm for extracting the signature (the distance between the centroid and contour). At this stage we apply morphological operators: dilation and erosion to obtain a complete pattern (without holes) and we calculate the perimeter and the area of the shape. Another parameter of the fuzzy system: the shape symmetry is determined based on the signature. The 3rd step includes a fuzzy logic decision support system which, based on fuzzy input variables (shape, size and symmetry), groups the packaging into three classes: recyclable, dangerous and determined. At the fourth step based on a difference function of signatures, the decision taken at the previous step is confirmed or refuted.

2 Automated identification of objects based on NCC and GA

Dangerous symbol identification on the label is done by templates matching with the perilous symbols [7] using CCN. The genetic algorithm generates new values for rescale the image from 1/3 to 3 times and rotation from 0 to 360, so the results do not depend on the size and image capture angle. The method is described in detail in [6].

The implementation of a GA reduces the calculation time for new values of size and rotation angle of the image, preserving only the best previous results at generation of new values. GA reduces the calculation steps thus achieving an optimal outcome time.

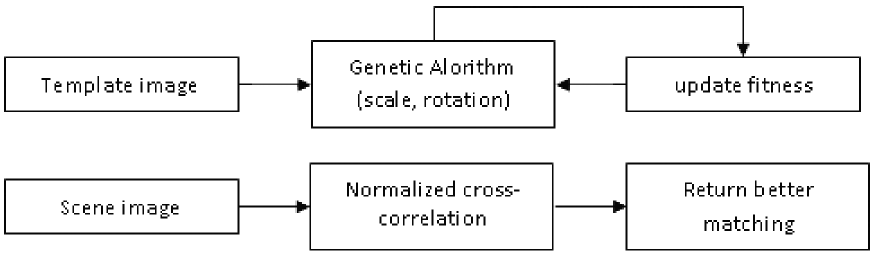


Figure 1. The block diagram of the method based on normalized cross-correlation and GA [6]

Table 1. Implementation of the method based on CCN and GA

Template and packaging picture	Proposed method implementation	Final result
		

The benefits of implementing AG are argued and presented in [6].

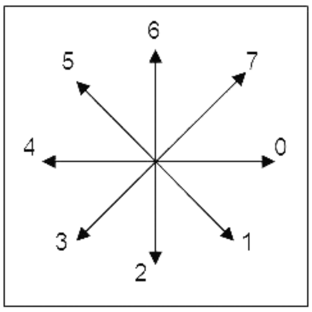
3 Features extraction

The process of the shape feature extraction comprises the following steps:

- image binarization, segmentation using a threshold histogram method;
- application of morphological operators: dilation, erosion;
- implementation of a contour tracking algorithm;
- implementation of a signature extraction algorithm;
- determination of shape, size, symmetry.

3.1 Contours determination

For image binarization it is used a threshold algorithm [8] or [9] with a single threshold. Having the binary image, we are looking for selection of the contour pixels, hence the name of the algorithm – contour tracking algorithm [10]. Contours tracking algorithm (walk of the blind) scans the image from the upper left corner until it finds a pixel (different from the background) which belongs to the object, this pixel is denoted by Ps and it is the start pixel of the contour. The current pixel is denoted by Pc. It is chosen a scroll direction of the contour. It is defined a variable dir that holds previous moving direction of the previous item along the contour to the current one. It initializes $dir = 0$ and $arm = (dir + 5)$ modulo 8.

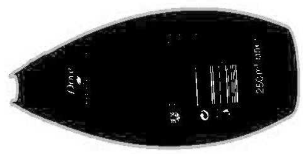
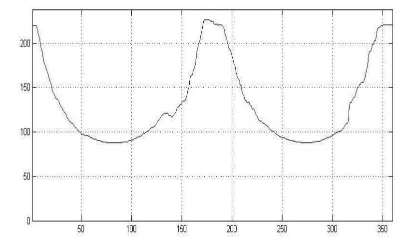
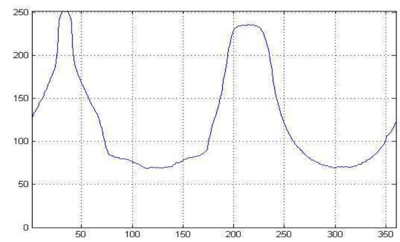


$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j+1}$
$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

Figure 2. Representation of the direction and vicinity 3x3 of the current pixel

It scans the neighbors of current pixel in the chosen scroll direction starting from the arm position, until finding a new pixel object which is labeled and becomes current pixel Pc. Variable direction dir is updated based on the current pixel found, and the variable arm is recalculated according to dir. The previous step is repeated until the closure of contour, then until the start pixel becomes again the current pixel. Examples of contour determination are shown in Table 2, column 1.

Table 2. Edge and signatures representation

Edge determination	Shape signatures
	
	

3.2 Shape-signature extraction

A signature is a one-dimensional representation of a surface. It was chosen to define the signature as the distance from a reference point – the gravity center of shape (g_x, g_y) – to every point of the contour (Table 2, column 2) [11]. The idea is to reduce the representation of contour to a one-dimensional function, which is basically easier to describe than the original two-dimensional outline.

The shape's signature is a reversible representation – knowing the signature it is possible to reconstruct the object contour. The size of the signature vector is of 360 values, one value obtained for each angle value of half-straight line which starts from the point (g_x, g_y) and intersects the contour (the edge). If the intersection of a half-straight line with the contour which scans the shape is made at several points, then it will preserve the farthest value.

$$d(n) = \sqrt{(x(n) - g_x)^2 + (y(n) - g_y)^2}. \tag{1}$$

3.3 Shape, size, symmetry determination

The paper [12] is a review of description and representation techniques of forms in order to recognize the objects or their automatic classification. The number of distinctive features is directly proportional to the correctness of the obtained results after execution of an Pattern Recognition algorithm and indirect to computing time. For example, if the color is not a distinctive feature, it is not recommended to be included in the algorithm. For packages description there were chosen three distinctive characteristics: shape, size and symmetry. Before calculating the area of the object, the morphological operators were applied: dilatation – to fill "holes" which were obtained from segmentation due to light reflection, and then erosion – to return to the original shape (delete pixels that were added at dilatation). The area (A) is considered to be the total number of pixels that represent the object. The perimeter (P) is represented by the pixels of contour. The shape and the size are calculated according to the formulas below:

$$Shape = \frac{4\pi A}{P^2}, \tag{2}$$

$$Size = \frac{A\_shape}{A\_image}. \tag{3}$$

A_shape represents the area of object that we want to identify and the *A_image* – is the total number of pixels from image. The contour of shape is not a smooth line after segmentation, but contains unevenness, therefore at the application of the formula for form calculation, we get approximate results, with a margin of error. For this reason it is done a normalization so that the values should be in the range [0, 1]. The symmetry is determined by signature. We assume that the symmetrical shapes (after horizontal axis) represent dangerous packaging – bottles of spray under pressure, batteries etc.

4 Fuzzy Rule-based System

Fuzzy logic differs from binary logic based on variables 0 or 1 ("true" or "false") by intermediate values, which is "true" or "false" with a degree of membership, 0 representing not belonging and 1 – total belonging. The fuzzy system transforms linguistic expressions "huge size", "elongated shape", "small symmetry" that include imprecision / ambiguity in numerical values. Hence, the representing degrees of membership in interval [0, 1], correspond to the partial truth between "completely true" and "completely false". There can be used modifiers / linguistic enhancers "somewhat high", "little", "pretty little", "extremely" etc. with numerical values for the system inputs. The Fuzzy logic renders values.

1) **The definition of fuzzy sets for input variables**

The fuzzification of inputs consists in determining the belonging degree of input data to fuzzy sets, by the membership function. The most usual membership functions are triangular and trapezoidal shape. We have chosen the trapezoidal membership function to define the set, because compared to the triangular membership functions, it has an interval in which the membership degree is 1, but the choice is subjective.

a) Definition of fuzzy sets for input variable *Size*

The input variable *Size* is defined by four trapezoidal fuzzy sets: Small, Medium, Large, and Extra Large with the universe of discourse [0, 0.5]. It is considered that the huge-sized objects can occupy up to 50% of the viewing area, hence the upper limit of the universe of discourse is 0, 5. A large size object covers about 10%–15% of the viewing area. According to Figure 3, when an object occupies between 15% and 20% of the viewing area, then it can be considered as belonging also to the fuzzy set "Large" and the fuzzy set "Extra Large", but with different degrees of belonging. The 4 fuzzy sets for input variable *Size* are represented in Figure 3.

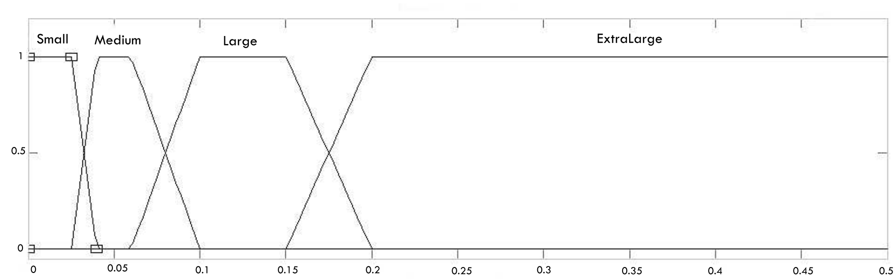


Figure 3. The fuzzy sets for input variable *Size*

b) Definition of fuzzy sets for input variable *Shape*

The input variable *Shape* is defined by four trapezoidal fuzzy sets: Elongated, Irregular, Oval and Round. According to the formula (2) the obtained values for *Shape* can be in the range [0, 1], with 0 corresponding to an elongate object similar to a line segment and 1 – to a circle. The fuzzy sets for input variable *Shape* are represented in Figure 4.

c) Definition of fuzzy sets for input variable *Symmetry*

The input variable *Symmetry* is defined by three trapezoidal fuzzy sets: Small, Medium and Big. It is expressed dimensionless based on differences between the normalized signature of the object on the intervals [0, 90], [90, 180], [180, 270], [270, 360] degrees. The definition of fuzzy sets for *Symmetry* was based on our observation: the bottles under pressure (from various sprays, lacquers), the batteries are sym-

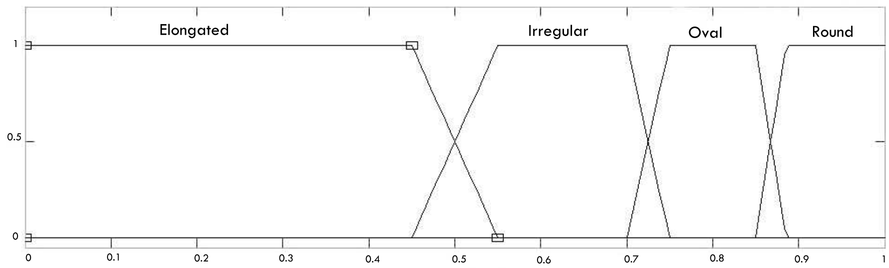


Figure 4. The fuzzy sets for input variable *Shape*

metrical towards the horizontal axis too. These packages / objects are dangerous and can't be recycled, so the membership on the Symmetry set is an important weight in decision made by the system. The fuzzy sets for input variable *Symmetry* are represented in Figure 5.

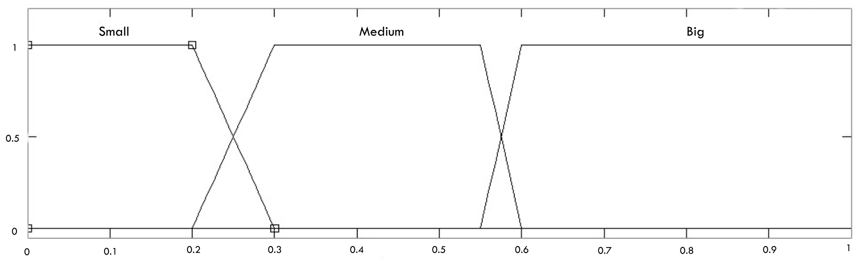


Figure 5. The fuzzy sets for input variable *Symmetry*

2) **The definition of fuzzy sets for the output variables**

Fuzzy system classifies waste into three classes: Dangerous (toxic, flammable, harmful to the environment), Undetermined and Recyclable. The output variable ObjectType is defined by fuzzy sets as in Figure 6, creating a Mamdani type system. Triangular membership functions have the small value of the triangle base (symbolical value), therefore the system can be considered closer to a Sugeno fuzzy system. Choosing the type of system Mamdani and not Sugeno is given by the defuzzification function MoM (Moment of Maximum) type. Small base values of the 3 fuzzy sets are chosen as follows: about 0 value for Dangerous, ~ 0.5 for Undetermined and ~ 1 for Recyclable.

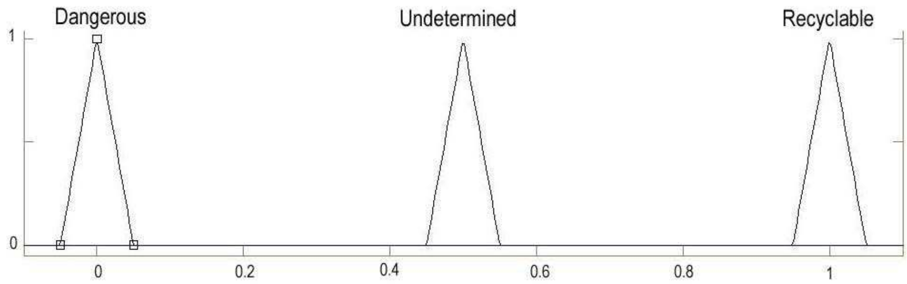


Figure 6. Fuzzy output representation

3) **Rules base of fuzzy logic system**

The rules of inference are the rules that bound fuzzy input variables of a system with fuzzy output variables of the same system. These rules are in the form as follows: IF condition / premise1 AND / OR condition / premise2 (AND / OR ...) THEN action. The entirety of rules forms the basis of rules, called Knowledge Base. The rules are qualitative statements derived from analyses made during recycling process of packaging. On this basis, an intuitive enunciation of fuzzy rules for the submitted application was defined: different weights associated to the rules can be observed. These weights are also called coefficients of certainty.

Examples of applied inference rules:

IF elongated shape AND small size AND big symmetry THEN object type is dangerous (1)

IF elongated shape AND medium size AND big symmetry THEN object type is dangerous (1)

IF elongated shape AND small size AND medium symmetry THEN object type is dangerous (0.7)

IF elongated shape AND medium size AND small symmetry THEN object type is undetermined (1)

IF elongated shape AND small size AND small symmetry THEN object type is undetermined (1)

IF elongated shape AND extra-large size AND big symmetry THEN object type is undetermined (1) etc.

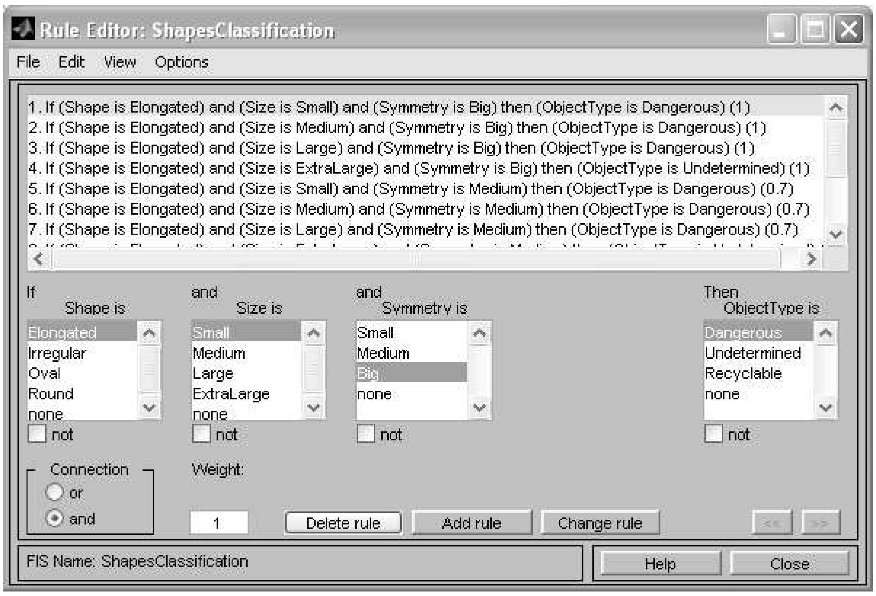


Figure 7. Inference rules applied to SF

The rules may be modified, supplemented, deleted via a graphical interface shown in Figure 7.

The intersection of two fuzzy sets is equivalent to the operation "AND logic" which is the minimum of two membership degrees. Between the premises of the rules the operator "min" is applied, the resulting value is used to truncation of output variable. When we have more active rules, a reunion of truncated fuzzy sets resulted from application of each rule is applied.

4) **Defuzzification**

In order that the developed system to classify Object Type, it is necessary that fuzzy sets of the output to be transformed into crisp data. This transformation is called defuzzification. The most common methods of defuzzification are: Centroid, COA, COG, SOM (smallest of maximum), MOM (mean of maximum), LOM (largest of maximum). For the developed system the **MoM** method was chosen. The fuzzy system for detection type of packaging applies the defuzzification operator MOM, selects from active fuzzy rules. An example of defuzzification is shown in Figure 8. Observe the active rules 7, 8 and 13. Applying MoM – the mean of maximums (of 0, 0,5 and 1), we get the result 0,5. If we have two activated rules, with the same weight, mean of 0 and 0,5 or mean of 0,5 and 1, in addition to relation MoM we apply the logical relationship AND – minimum of two values.

The achievement of this system is 95%. Some packages from the group Undetermined were identified as Dangerous. A Dangerous object has been classified as Undetermined. Most importantly, these packages are not classified as Recyclable, but to ensure an accurate automatic classification, the items from the Recyclable class are checked at the next step by identifying the signature (of the form).

5 Signatures identification

The forms signature is a functional representation. For the proposed system the signature representation by the distance from the form cen-

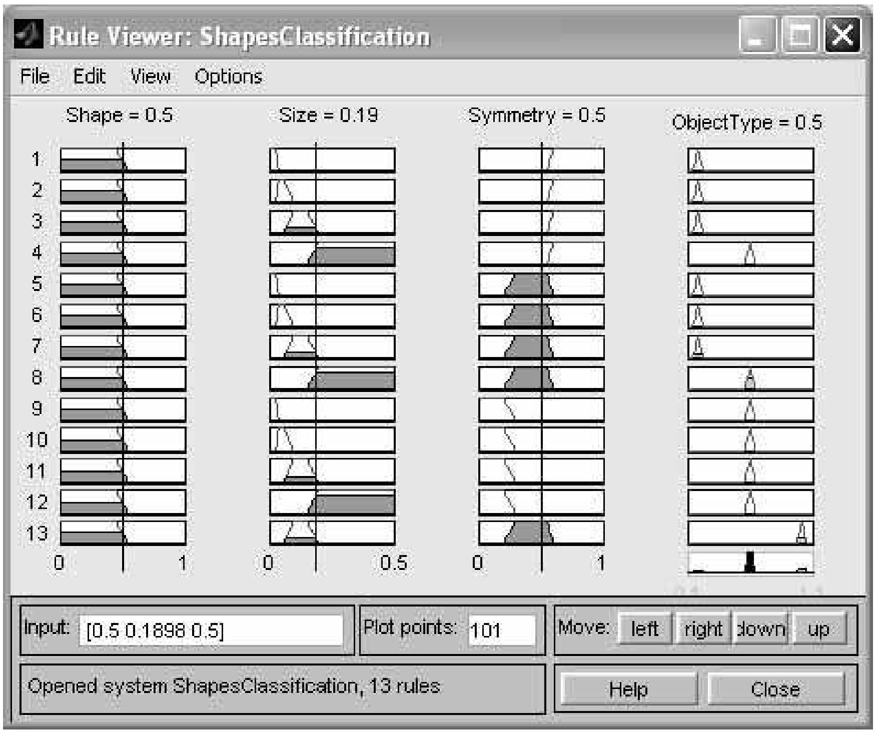
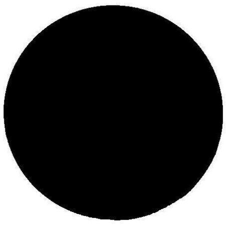
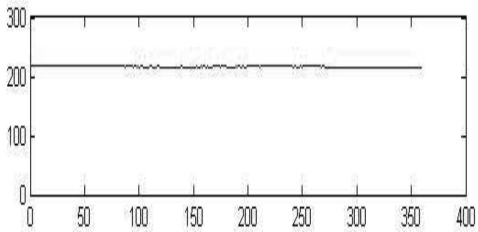
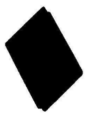
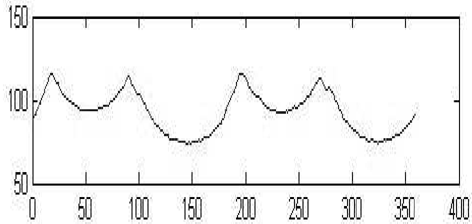


Figure 8. Example of defuzzification

ter to each point of edge was chosen. Examples of shapes and their signatures are shown in Table 3.

For objects with the same shape but different sizes, the signatures will be similar but with different amplitude (due to varying degrees of scaling). For objects with different orientation, the signatures will be similar but moved to the left / right. Applying normalization to the signature, the effect of scaling disappears. Both the correlation function and difference function ensure the invariance to rotation. Another possibility to eliminate the effect of rotation is selecting the starting point for signature generation as the farthest point from the center (if it is a single point) or the farthest point on the axis. One way to get the

Table 3. Graphic representations of forms signatures

<table>
<tr><th>Examples of forms</th><th>Signatures of these forms</th></tr>
<tr><td></td><td>
</td></tr>
<tr><td></td><td>
</td></tr>
</table>

invariance of forms size would be the normalization of obtained values: their representation between 0 and 1. This method is sensitive to noises that affect the maximum and minimum values. For the developed system the size of the packaging matters, therefore no normalization was performed. The correlation is a statistical method applied to determine a relationship between two or more variables. The correlation coefficient (r) is a quantitative value between -1 and $+1$, and describes the relationship between variables (see Table 4). Extreme values imply a correlation between variables while $r = 0$ is the total lack of linear relationship.

Examples of signatures identification by total correlation coefficient are represented in Table 5.

Defined for two variables with normal distribution (x, y), the lin-

Table 4. Guidelines to interpreting Pearson's correlation coefficient

Strength of association	r **positive**	r **negative**
Small	0.1 to 0.3	-0.1 to -0.3
Medium	0.3 to 0.5	-0.3 to -0.5
Large	0.5 to 1.0	-0.5 to -1.0

ear correlation coefficient called Pearson's coefficient (total correlation coefficient), is determined as follows [13]:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y)}{\sqrt{(\sum_{i=1}^{n}(x_i - m_x)^2(\sum_{i=1}^{n}(y_i - m_y)^2}}, \tag{4}$$

where

m_x – the average of values $x : m_x = \sum_{i=1}^{n} \frac{x_i}{n}$,

m_y – the average of values $y : m_y = \sum_{i=1}^{n} \frac{y_i}{n}$.

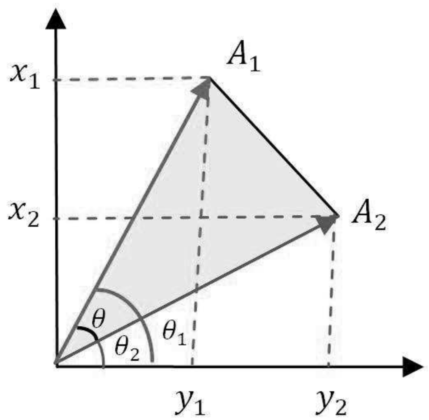
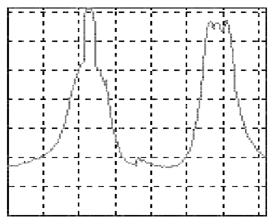
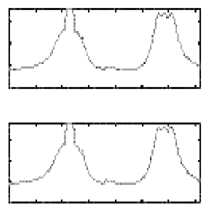
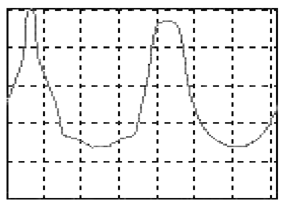
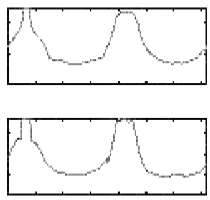
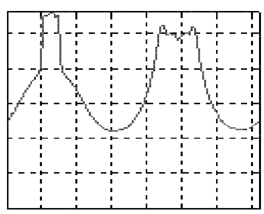
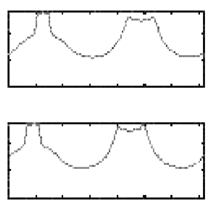


Figure 9. The cosine θ coefficient for a two-dimensional space [13]

Calculation of cosine θ coefficient (see Figure 9) in a two-dimensional space is based on trigonometric relationships of angles and is expressed

Table 5. Signatures identification by total correlation coefficient

Object shape	Shape signature	Shape signature identification
 Recyclable item, included in the DB		 Recyclable (pet10.bmp)
 Recyclable item, not included in the DB		 Recyclable (pet12.bmp)
 Unrecycled item intentionally included		 Undetermined (ambalaj.bmp)

by the following formula:

$$cos\theta_{A_1A_2} = \cos(\theta_1 - \theta_2) = \frac{x_1y_1 + x_2y_2}{\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}}. \quad (5)$$

The above formula can be adapted for n independent factors:

$$cos\theta_{A_1A_2} = \frac{\sum_{i=1}^{n} x_iy_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}. \quad (6)$$

The cosine θ coefficient is a measure of angular distance used for estimating similarity between the forms. This correlation coefficient indicates a complete similarity between the two forms A_1 and A_2 if $\cos\theta = 1$ and a total distinction if $\cos\theta = 0$. The both formulas were tested and we opted for the total correlation coefficient (called Pearson's coefficient). The database contains the signatures of all forms (recyclable undetermined and dangerous) to be identified as a non-recyclable object if it has been misclassified. The objects that are not included in the database were also tested and, the result is satisfactory.

6 Conclusions

The elaborated decision system aims at detection of dangerous waste that appears because of ignorance of people with respect to recyclable waste.

The Hybrid algorithm – CCN combined with GA – enables the detection of danger symbols on the label with a very high rate of success. Implementation of genetic algorithm for generating rotation and scaling values is a solution when template differs in size and rotation angle of the image scenes. GA also reduces calculation steps reaching an optimal result within a reasonable time.

Objects contour extracting in order to obtain the signature is made by the tracking contour algorithm. Correct determination of the contour also enables efficient calculation of shape and size – features used in fuzzy system.

Automatic classification of waste is carried out with a fuzzy system the inputs of which are: shape, size and symmetry. The choice of fuzzy sets is subjective and was made after analysis of packaging that should be recycled, and the dangerous ones must be collected in special containers. To increase the weight of decision that some waste is dangerous or can be recycled, an algorithm for form (packaging) signature identification is implemented.

The developed system can be adapted to sort other objects. Waste sorting was chosen to emphasize that their correct sorting would be a solution for reducing pollution, an important factor nowadays.

References

[1] M.S. Bani, Z.A. Rashid, K.H.K. Hamid, M.E. Harbawi, A.B. Alias and M.J. Aris, "The Development of Decision Support System for Waste Management; a Review," *Int. Journal of Chemical, Molecular, Nuclear, Materials and Metallurgical Engineering*, vol. 3, no. 1, pp.17–24, Jan. 2009.

[2] S. Cheng, C.W. Chan and G.H. Huang, "An integrated multi-criteria decision analysis and inexact mixed integer linear programming approach for solid waste management," *Engineering Application of Artificial Intelligence*, vol. 16, pp. 543–554, Aug.-Sept. 2003.

[3] N. Musee, L. Lorenzen and C. Aldrich, "New methodology for hazardous waste classification using fuzzy set theory," *Journal of Hazardous Materials*, vol. 154, pp. 1040–1051, June 2008.

[4] H. Stirling, "The recovery of waste glass cullet for recycling purposes by means of electro-optical sorters," *Conservation and Recycling*, vol. 1, no. 2, pp. 209–219, Dec. 1977.

[5] M.O. Rahman, A. Hussain, E. Scavino, H. Basri and M.A. Hannan, "Intelligent computer vision system for segregating recyclable waste papers," *Expert Systems with Applications*, vol. 38, pp. 10398–10407, Aug. 2011.

[6] M. Rusu and M.-D. Zbancioc, "Automated identification of objects based on Normalized Cross-Correlation and Genetic Algorithm," in *IEEE Int. Conf. on E-Health and Bioengineering*, 19-21 Nov. 2015, pp. 1–4.

[7] "Annex 1: Symbols and indications of danger", *International labour Organization*. [Online]. Available: http://www.ilo.org/legacy/english/protection/safework/cis/ products/safetytm/clasann1.htm. Accessed on Sept. 30, 2015.

[8] M. Rusu and H.N. Teodorescu, "A Method for Image Segmentation based on Histograms – Preliminary Results," in *The 4-th Int. Conf. on Telecommunications, Electronics and Informatics*, 17-20 May 2012, pp. 351–354.

[9] N. Otsu, "A threshold selection method from graylevel histograms," in *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, Jan. 1979, pp. 62–66.

[10] Chapter 5.2 "Border tracing", Section from Chapter 5, part II *Segmentation: Edge-based segmentation*. [Online]. Available: http://user.engineering.uiowa.edu/∼dip/LECTURE/Segmentati on2.html. Accessed on: Feb. 02, 2016.

[11] M. Yang, K. Kpalma and J. Ronsin, "Shape-Based Invariant Feature Extraction for Object Recognition," in *Advances in Reasoning-Based Image Processing Intelligent Systems*, (Intelligent Systems Reference Library, vol. 29), R. Kountchev and K.

Nakamatsu, Eds. Springer Berlin Heidelberg, Jan. 2012, pp. 255–314.

[12] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 119, Jan. 2004.

[13] D. Scradeanu, *Modele cantitative statistice*, 104 p. [Online]. Available: http://www.unibuc.ro/prof/scradeanu_d/docs/2013/noi/08_21_27_123_Modele_CANTITATIVE_STATISTICE_DSCRD.pdf.

[14] K. Jimoh, A. Ajayi and O. Ayilara, "Intelligent Model for Manual Sorting of Plastic Wastes," *Int. Journal of Computer Applications*, vol. 101, No. 7, pp. 20–27, Sept. 2014.

Mariana Rusu, Marius-Dan Zbancioc
Received October 27, 2016

Mariana Rusu
Technical University of Moldova
Str. Studentilor 9/7, MD-2045, Chisinau, Republic of Moldova
E–mail: `mariana.rusu@ffi.utm.md`

Marius-Dan Zbancioc
1. "Gheorghe Asachi" Technical University of Iasi
Romania, Iasi, 8 Bd. Carol I, and
2. Institute of Computer Science of the Romanian Academy
Romania, Iasi
E–mail: `zmarius@etti.tuiasi.ro`