rough sets,
cellular automata,
medical image processing

Bartłomiej PŁACZEK[1]

ROUGH SETS IN IDENTIFICATION OF CELLULAR AUTOMATA FOR MEDICAL IMAGE PROCESSING

In this paper a method is proposed which enables identification of cellular automata (CA) that extract low-level features in medical images. The CA identification problem includes determination of neighbourhood and transition rule on the basis of training images. The proposed solution uses data mining techniques based on rough sets theory. Neighbourhood is detected by reducts calculations and rule-learning algorithms are applied to induce transition rules for CA. Experiments were performed to explore the possibility of CA identification for boundary detection, convex hull transformation and skeletonization of binary images. The experimental results show that the proposed approach allows finding CA rules that are useful for extraction of specific features in microscopic images of blood specimens.

1. INTRODUCTION

Cellular automata (CA) are discrete dynamic systems composed of an array of cells, where each cell can be in one of a finite number of possible states. The state of cells is updated synchronously in discrete time steps according to local transition rules (CA rules). The local transition rules are used to compute state of a cell at the next time step by taking into account current state of cells in a predefined neighbourhood.

Recent medical applications of CA include: simulation of tumour growth [3], information extraction from clinical reports [5], data encryption in telemedicine systems [4] and processing of medical images [10]. Due to the discrete structure and local, parallel computations, CA appear as appropriate and useful tools for digital image processing. In this field, CA were successfully applied to edge detection, noise reduction, morphological transformations, image segmentation, and compression tasks [13], [16].

While numerous works in the literature have been devoted to CA applications, relatively little research effort has been directed towards automatic identification of CA, i.e. algorithmic determination of rules (and associated neighbourhood) that allow the CA system to achieve a desired result. For most of the existing applications, the update rules and neighbourhood for CA have been determined by human experts. Automatic and computationally efficient identification of the CA for image processing systems remains an open research issue.

This paper discusses the possibility of using data mining techniques based on rough sets theory [6] to determine rules of CA for feature extraction in medical images. According to the proposed approach, input data describing pixel values are represented in form of a decision table (attribute-value system) and a rule-learning method is applied to induce a set of decision rules that define the evolution of CA. Neighbourhood is detected by using algorithms for reducts calculation. This approach was used for

[1]Institute of Computer Science, University of Silesia, Będzińska 39, 41-200 Sosnowiec, Poland
placzek.bartlomiej@gmail.com.

identification of CA that perform boundary detection, convex hull transformation and skeletonization of binary images. The experimental results show that the proposed approach allows finding CA rules that are useful for extraction of specific features in medical images.

The remainder of this paper is organised as follows. Related works are reviewed in Section 2. Section 3 discusses the problem of CA identification for image processing applications. The proposed rough sets approach to the identification problem is presented in Section 4. Section 5 includes experimental result of feature extraction in microscopic blood images. Finally, conclusions are given in Section 6.

2. RELATED WORKS

Most of the works related to the CA identification problem use evolutionary algorithms as a tool to extract update rule and neighbourhood from spatio-temporal patterns produced in CA evolution [7], [12], [20]. This approach was applied for different classes of CA, e.g. in [15] a genetic algorithm was employed to learn probabilistic rules directly from experimental data for two-dimensional binary-state CA.

Although application of the evolutionary algorithms was a dominant approach to CA identification, some non-genetic techniques are also available. Adamatzky [1], [2] proposed several approaches to extracting rules for different classes of CA. Straatman et al. [18] developed a form of hill climbing and backtracking to identify CA rules. In [16] a deterministic sequential floating forward search method was used to select rules of CA. Another approach is based on parameter estimation methods from the field of system identification [21]. A framework for solving the identification task for both deterministic and probabilistic CA was presented in [19].

A number of works in the literature have addressed the problem of rules generation for the CA-based image processing systems. Most of the authors have used the evolutionary algorithms, e.g. in [17] this approach was applied for the identification of CA that perform edge detection operations. Another example is the method of hyper-spectral images segmentation through the use of CA [14], where the rule set is automatically obtained by using an evolutionary algorithm. Several methods were proposed that use genetic algorithms to learn both CA rules and neighbourhood size [8], [9]. In [16] a deterministic sequential floating forward search method for feature selection was used to select effective CA rules for several image processing operations (filtering, thinning, and convex hulls).

The existing methods of CA identification for image processing tasks are based on iterative procedures, in which the CA with a set of candidate rules is applied repeatedly for training images and its performance is evaluated. At successive iterations, the set of candidate rules is modified according to a specific strategy, with taking into account the performance rating. Even if only binary images are considered, the space of all possible rule sets is very large, thus the iterative CA identification procedure requires long computational time [16].

In this paper a method is proposed, which finds the CA rules in a single iteration by using a rule-learning algorithm based on rough sets theory. The rules are induced from an information table, which includes results of a comparison between input and target images. This method was applied for feature extraction in microscopic images of blood specimens. The introductory research results shows that this approach enables a fast and effective selection of useful CA rules for the considered image processing tasks.

3. PROBLEM FORMULATION

A cellular automaton is a dynamic system, which is discrete in space and time and operates on an array of cells. Formally, a cellular automaton can be defined as a triple (S, N, δ), where S is a non-empty set, called the state set, N is the neighbourhood, and δ is the local transition rule. In this study two-dimensional deterministic CA are considered, thus $N \subseteq \mathbb{Z}^2$ and $\delta : S^{|N|} \mapsto S$. Arguments of δ are the current states of cells in the neighbourhood, while the value of δ indicates the state of a central cell at the next time step.

When CA are used for image processing, the cell states correspond to pixel values and S is a set of the allowable pixel values ($S = \{0, 1\}$ for binary images). The problem of CA identification involves finding both the cells neighbourhood N and the transition rule δ on the basis of a training data set, which includes an input image X and a target image Y. The identified cellular automaton has to enable transformation of the input image into the target image by using the transition rule.

Let us assume that at the first time step of the cellular automaton application ($t = 0$) pixel value $x_{i,j}(t)$ is determined by the input image X. At successive time steps ($t = 1, 2, \ldots, T$) new pixel value is calculated using the transition rule as follows:

$$x_{i,j}(t+1) = \delta\left(x_{g,h}(t), \ldots, x_{k,l}(t)\right), \tag{1}$$

where $\{(g,h), \ldots, (k,l)\} = N$.

Such operation is performed for all pixels at each time step. The procedure is finished after $T+1$ time steps, when the application of transition rule does not result in further changes of the pixel values. Thus, the pixel values $x_{i,j}(T+1)$ define an output image X'. The objective is to find such transition rule δ (and associated neighbourhood N), for which the output image X' will be identical or at least similar to the target image Y.

4. PROPOSED APPROACH

In order to use the rough sets approach for solving the problem defined in Sect. 3, the training set of data has to be represented in the form of a decision table $I = (U, C)$, where U is a non-empty set of observations and C is a set of pixel values (cell states):

$$C = \{x_{a,b}, \ldots, x_{i,j}, \ldots, x_{m,n}, y_{i,j}\} \tag{2}$$

where $x_{i,j}$ and $y_{i,j}$ are values of the pixel (i,j) in input and target image respectively.

The neighbourhood and transition rule have to be found for the central pixel (i,j), thus its value in the target image $y_{i,j}$ is used as the decision attribute. The pixel values in input image $x_{a,b}, \ldots, x_{m,n}$ are condition attributes. Pixels $(a,b), \ldots, (m,n)$ are considered as candidate neighbours of the central pixel (i,j). Observations registered in the decision table correspond to particular pixels (i,j) of the input image.

Neighbourhood for the pixel (i,j) can be selected by calculating reducts of the above-defined decision table. A reduct is a subset of the condition attributes which is sufficient to determine the decision attributes [6]. Taking into account the decision table discussed above, reduct should be defined as a subset of pixel values $R \subseteq \{x_{a,b}, \ldots, x_{m,n}\}$, which preserves discernibility of the observations with regard to the decision $y_{i,j}$, and none of its proper subsets has this ability. Observations are discernible if they differ in at least one condition attribute (pixel value). Each two observations that have different decisions $y_{i,j}$ and are discernible by the full set of pixel values $\{x_{a,b}, \ldots, x_{m,n}\}$ are also discernible by the reduct R.

The neighbourhood N of a pixel (cell) is determined as a set of coordinates of the pixels whose values belong to the reduct $R = \{x_{g,h}, \ldots, x_{k,l}\}$:

$$N = \{(g,h), \ldots, (k,l)\}. \tag{3}$$

There may exist multiple reducts for one decision table. Selection of the neighbourhood is made with regard to the shortest reduct, because size of the neighbourhood has to be minimized.

When the reduct R is found, the condition attributes that do not belong to this reduct are excluded from the decision table. Thus, the modified decision table $I' = (U, C')$ has the following set of attributes:

$$C' = R \cup y_{i,j}, \tag{4}$$

where $y_{i,j}$ remains the decision attribute.

Transition rule of a cellular automaton is identified as a set of decision rules by taking into account the information from the modified decision table I'. A particular decision rule r has the following form:

$$(x_{g,h} = \alpha) \wedge \ldots \wedge (x_{k,l} = \beta) \Rightarrow x_{i,j} = \gamma, \tag{5}$$

where $\alpha, \beta, \gamma \in S$, and S denotes the set of allowable pixel values (cell states). From a set of the induced decision rules only those are selected that may change the value of the central pixel (i, j).

Two characteristics of the decision rules are useful for the proposed method: support and match [6]. Support of a rule r, denoted by $SUPP_{I'}(r)$, is equal to the number of observations from I' for which the rule r applies correctly, i.e., premise of the rule is satisfied and the decision given by rule is consistent with the one in decision table. $MATCH_{I'}(r)$ is the number of observations in I' for which the rule r applies, i.e., premise of the rule is satisfied. Based on these two characteristics, a certainty factor is defined for the decision rule r:

$$CER_{I'}(r) = SUPP_{I'}(r)/MATCH_{I'}(r). \tag{6}$$

The certainty factor may be interpreted as a conditional probability that decision of rule r is consistent with an observation in the decision table I', given that premise of the rule is satisfied. In the proposed method, this factor is taken into account to discard uncertain decision rules.

Fig. 1 summarizes the main operations that are necessary to identify a cellular automaton for image processing by using the rough sets approach. In this study, reducts and decision rules are calculated using algorithms implemented in the RSES software [6]. Three algorithms of reducts calculation were examined: exhaustive, genetic, and dynamic reduct algorithm. Moreover, the experiments involved application of three algorithms that enable induction of decision rules: exhaustive, genetic, and LEM2 [11].

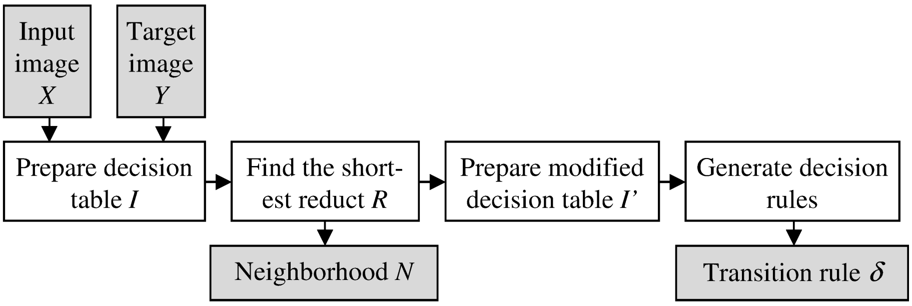

Fig. 1. Rough sets based procedure of CA identification.

5. EXPERIMENTS

In this section results are presented of applying the proposed approach for identification of CA that perform feature extraction in medical images. The considered application examples include: boundary detection, convex hull transformation, and skeletonization. Tests of the identified CA were performed by using binarized microscopic images of blood specimens. For the presentation of binary images it was assumed that pixels with value 1 are black while 0-valued pixels are white.

For all examples discussed here the neighbourhood N of a pixel (i, j) contains the pixel itself and its eight nearest neighbours (see Fig. 2 c). The neighbourhood was recognized as the shortest reduct by each of the reducts calculation algorithms (exhaustive, genetic, and dynamic). Thus, the decision table I' for each example includes ten attributes:

$$C' = \{x_{i-1,j-1}, x_{i,j-1}, x_{i+1,j-1}, x_{i-1,j}, x_{i,j}, x_{i+1,j}, x_{i-1,j+1}, x_{i,j+1}, x_{i+1,j+1}, y_{i,j}\}. \tag{7}$$

Hereinafter, the decision rules are presented in figures by using 3 x 3 pixel masks where black and white colours indicate pixels with value 0 and 1 respectively. A particular decision rule applies if the corresponding mask is consistent with local pixel configuration in a processed image.

An example of boundary detection for a blood cells image is illustrated in Fig. 2. Input and target images that were used for the identification of CA are shown in Fig. 2 a-b. On this basis a set of the decision rules was induced. Only one rule was found, which changes value of the central pixel. A pixel mask for this rule is illustrated in Fig. 2 c (note that the black colour in figures indicates pixels with value 1). If a configuration of pixels in a processed image is consistent with the mask then the central pixel takes value 0. Thus, the decision rule can be written as follows:

$$\begin{aligned}&(x_{i-1,j-1}(t)=1)\wedge(x_{i,j-1}(t)=1)\wedge(x_{i+1,j-1}(t)=1)\\&\wedge(x_{i-1,j}(t)=1)\wedge(x_{i,j}(t)=1)\wedge(x_{i+1,j}(t)=1)\\&\wedge(x_{i-1,j+1}(t)=1)\wedge(x_{i,j+1}(t)=1)\wedge(x_{i+1,j+1}(t)=1)\Rightarrow x_{i,j}(t+1)=0,\end{aligned}\tag{8}$$

Application of CA with the above rule allows us to transform the input image (Fig. 2 a) into an output image, which is identical to the target image (Fig. 2 b). Fig 2 f shows results of using the identified CA for boundary detection in a test image (Fig. 2 e). The test image before binarization is presented in Fig. 2 d.

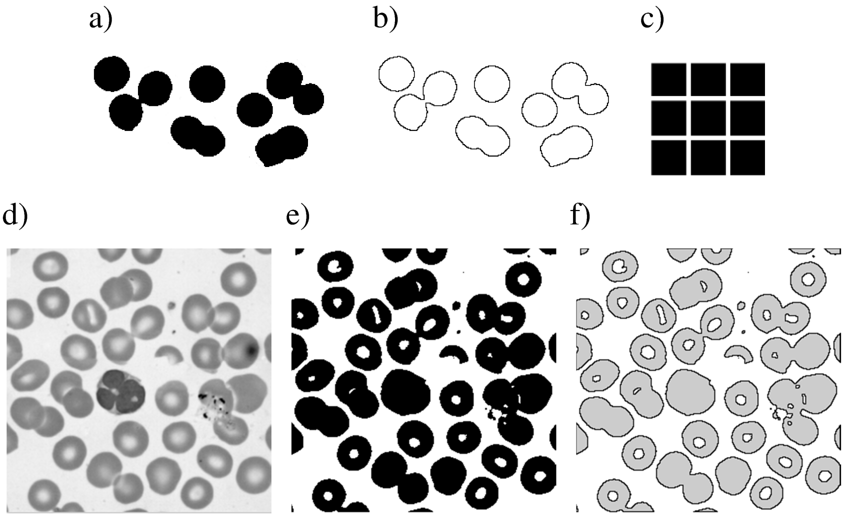

Fig. 2. Boundary detection: a) input image for training, b) target image, c) induced rule, d) original test image, e) binarized test image, f) results of CA application.

Results of CA identification for the convex hull transform are presented in Fig. 3. Fig 3 d shows a set of 8 certain decision rules induced on the basis of the training images (Fig. 3 a-b) by using algorithm LEM2. The decision rules are represented by pixel masks, where grey colour indicates pixels that can have an arbitrary value (the pixels are not taken into account by the decision rule). If pixels configuration in a processed image match one of the masks then value of the central pixel is changed to 1. The identified CA was used to transform the test image form Fig. 2 e. Results of this operation are illustrated in Fig. 3 e. The remaining algorithms of rule induction (exhaustive and genetic) have generated sets of decision rules that give identical results, however their size is larger (more than 8 certain rules). Note that the results in Fig. 3 were obtained by using only the certain decision rules (Fig. 3 d), i.e. those that have certainty factor $CER_{I'}(r) = 1$. In the next experiments, the set of decision rules was extended by adding uncertain rules, for which the certainty factor is above 0.6 (Fig. 3 f) and above 0.5 (Fig. 3 g). These tests show that the uncertain rules can significantly influence the results of an image processing operation.

The last two examples are devoted to skeletonization of a blood smear image with trypanosoma forms among red blood cells (Fig. 4 e). Fig. 4 shows the example, where CA was trained by using a target image (Fig 4 b), which was obtained by performing skeletonization as the standard morphological operation. The set of induced certain decision rules includes 9 elements (Fig. 4 d). During CA application, value of the central pixel is changed to 0 if the 3x3 pixel pattern is consistent with one of the nine masks presented in Fig. 4 d. In opposite situation the pixel value remains unchanged. Result of the CA-based skeletonization is shown in Fig 4 g (black pixels).

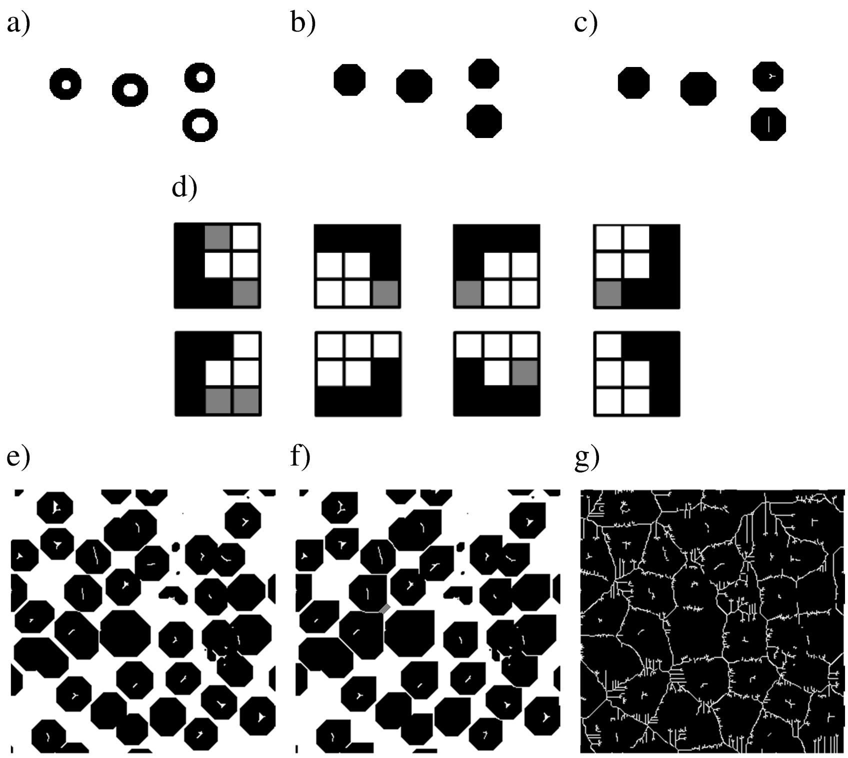

Fig. 3. Convex hull: a) input image for training, b) target image,c) results of CA application for input image, d) induced rules, e) - f) results of CA application for test image ($CER = 1$, $CER > 0.6$, $CER > 0.5$).

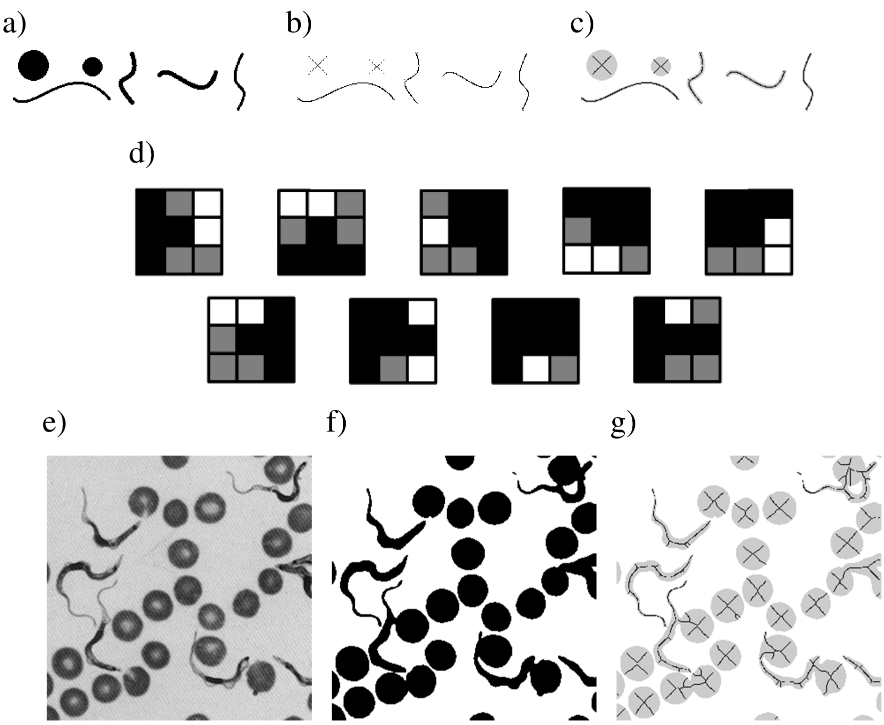

Fig. 4. Skeletonization: a) input image for training, b) target image, c) results of CA application for input image, d) induced rules, e) original test image, f) binarized test image, g) results of CA application for test image.

Figure 5 includes an example of custom skeletonization, where the skeletons of blood cells are removed as we are interested in recognizing skeletons of the trypanosoma forms. The input image for CA training (Fig. 5 b) was obtained by removing the X-shaped skeletons of circles from the image shown in Fig. 4 b. Results of the CA application with the induced rules are presented in Fig. 5 d (note that grey colour corresponds to the binary test image). It can be observed in this example that the CA can be used to detect the specific shape of trypanosoma forms in microscopic blood images.

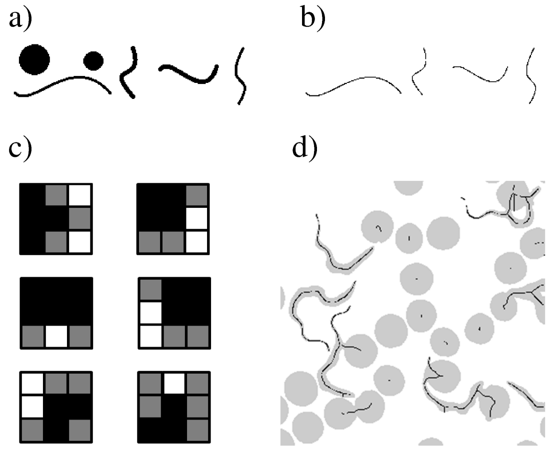

Fig. 5. Custom skeletonization: a) input image for training, b) target image, c) induced rules, d) results of CA application.

6. CONCLUSION

The data exploration techniques based on rough sets theory enable straightforward identification of CA for image processing on the basis of training data. The proposed method determines the neighbourhood and transition rule without iterative testing and modification of candidate solutions. Thus, the solution can be found in shorter time than for the evolutionary algorithms. The experimental results show that the pro-posed approach allows identification of CA that are useful for extraction of specific features in medical images.

In this study, experimental results obtained for the proposed method were evaluated by visual inspection. Detailed assessment of the introduced approach will require a definition of quality measures, a larger set of test images as well as comparisons against the existing methods. An interesting topic for further research is a possibility of combining the rough sets and evolutionary approaches in a hybrid system, where the rough sets method will be used for fast identification of a CA and an evolutionary algorithm will improve the obtained solution.

BIBLIOGRAPHY

[1] ADAMATZKY A., Automatic programming of cellular automata: identification approach. Kybernetes: The Int. J. of Systems & Cybernetics, 1997, Vol. 26, No. 2, pp. 126-135.

[2] ADAMATZKY A., Identification of Cellular Automata, Taylor & Francis, London, 1994.

[3] ALARCÓN T., BYRNE H. M., MAINI P. K., A cellular automaton model for tumour growth in inhomogeneous environment, Journal of Theoretical Biology, 2003, 225(2), pp. 257-274.

[4] ANGHELESCU P., IONITA S., SOFRON E., Encryption Technique with Programmable Cellular Automata (ETPCA). Journal of Cellular Automata, 2010, 5(1-2), pp. 79-105.

[5] BARIGOU F., ATMANI B., BELDJILALI B., Using a Cellular Automaton to Extract Medical Information from Clinical Reports. Journal of Information Processing Systems, 2012, 8(1), pp. 67-84.

[6] BAZAN J. G., SZCZUKA M., The rough set exploration system. In: J.F. Peters and A. Skowron (Eds.): Transactions on Rough Sets III, LNCS 3400, Springer, Heidelberg, 2005, pp. 37-56.

[7] BILLINGS S., YANG Y., Identification of probabilistic cellular automata, IEEE Trans. on Systems Man and Cybernetics, Part B: Cybernetics, 2003, Vol. 33, No. 2, pp. 225-236.

[8] CHAVOYA A., DUTHEN Y., Using a genetic algorithm to evolve cellular automata for 2D/3D computational development. In: Genetic and Evolutionary Computation Conf., 2006, pp. 231-232.

[9] CRAIU R. V., LEE T. C. M., Pattern generation using likelihood inference for cellular automata, IEEE Transactions on Image Processing, 2006, 15 (7), pp. 1718-1727.

[10] GHOSH P., ANTANI S. K., LONG L. R., THOMA G. R., Unsupervised Grow-Cut: Cellular Automata-based Medical Image Segmentation. First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology, 2011.

[11] GRZYMALA-BUSSE, J. W., LERS-a system for learning from examples based on rough sets. In: Intelligent decision support. Springer Netherlands, 1992, pp. 3-18.

[12] MAEDA K. I., SAKAMA C.: Identifying cellular automata rules. Journal of Cellular Automata, 2007, Vol. 2, No. 1, pp. 1-20.

[13] PEER M. A., QADIR F., KHAN K. A., Investigations of cellular automata game of life rules for noise filtering and edge detection. International Journal of Information Engineering and Electronic Business, 2012, 4(2), pp. 22-28.

[14] PRIEGO B., SOUTO D., BELLAS F., DURO R. J., Hyperspectral Image Segmentation through Evolved Cellular Automata, Pattern Recognition Letters, 2013, Vol. 34, No. 14, pp. 1648-1658.

[15] RICHARDS F. C., MEYER T. P., PACKARD N. H., Extracting cellular automaton rules directly from experimental data, Phys. D, 1990, Vol. 45, No. 1-3, pp. 189–202.

[16] ROSIN P., Training cellular automata for image processing, IEEE Trans. on Image Pro-cessing, 2006, Vol. 15, No. 7, pp. 2076-2087.

[17] SLATNIA S., BATOUCHE M., MELKEMI K. E., Evolutionary cellular automata based-approach for edge detection. In: Applications of Fuzzy Sets Theory, Springer, Berlin Heidelberg, 2007, pp. 404-411.

[18] STRAATMAN B., WHITE R., ENGELEN G., Towards an automatic calibration procedure for constrained cellular automata, Comput., Environ. Urban Syst., 2004, 28 (1-2), pp. 149-170.

[19] SUN X., ROSIN P. L., MARTIN R. R.: Fast rule identification and neighborhood selection for cellular automata. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 2011, 41 (3), pp. 749-760.

[20] YANG Y., BILLINGS S., Extracting Boolean rules from CA patterns, IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 2000, 30 (4), pp. 573-580.

[21] ZHAO Y., BILLINGS S., The identification of cellular automata, Journal of Cellular Automata, 2007, Vol. 2, No. 1, pp. 47-65.