

The two-step method of data mining from digital images

Anna Wójcicka^{1,2}, Roman Simiński², Zygmunt Wróbel²

¹Department of Technology and Engineering of Material, Pedagogical University, Institute of Technology, Podchorążych 2 St, 30-084 Cracow, Poland
aniawojcicka@gmail.com

²Department of Computer Biomedical Systems, University of Silesia, Institute of Computer Science, Będzińska 39 St, 41-200 Sosnowiec, Poland
roman.siminski@us.edu.pl, zygmunt.wrobel@us.edu.pl

Analysis of a coherent series of images might allow extraction from them interesting qualitative and quantitative features. Their further exploratory analysis may allow for the discovery occurring regularity, generalizations, connections and relationships [1]. Data mining is currently a field of research providing significant results in theoretical and practical aspects. The paper presents the problem of integration methods of image analysis and data mining using the two-step method of data mining from digital images which is developed by the authors. This method implemented in a computer system. The proposed method has to maximize knowledge acquisition automation capabilities from the images, while allowing for the use of knowledge and competence domain experts.

The described method data mining from digital images is divided into two stages. The schema of two-step method of data mining from digital images shown in figure 1.

The first step uses a selected method of image analysis [2, 3, 4, 5], focused on the extraction of quantitative and qualitative features from images. At this stage:

- determined number, type, name, and the names of individual characteristics or traits ranges,
- establish a set of graphical transformations, which are subjected to images in order to standardization and to obtain the required characteristics.

The result of this stage are tables with data, which are information system [6]. These data are subjected to pre-processing, which comprises the processing of the missing data, outliers and discretization. This pre-processed data has create the decision table [7]. In this table indicated the attribute decision and the conditional attributes. The decision table is input for the second step of the proposed method. This stage involves data mining which result to generate decision rules. Data mining is based on rough set theory [6]. The second stage of the method used to support domain experts. The experts using a dedicated system can verify the results based on the input images.

The last element of the method is the ability to use created knowledge base for the implementation of subject decision support system which has inference forward and backward module. The inference can be used for experimental and

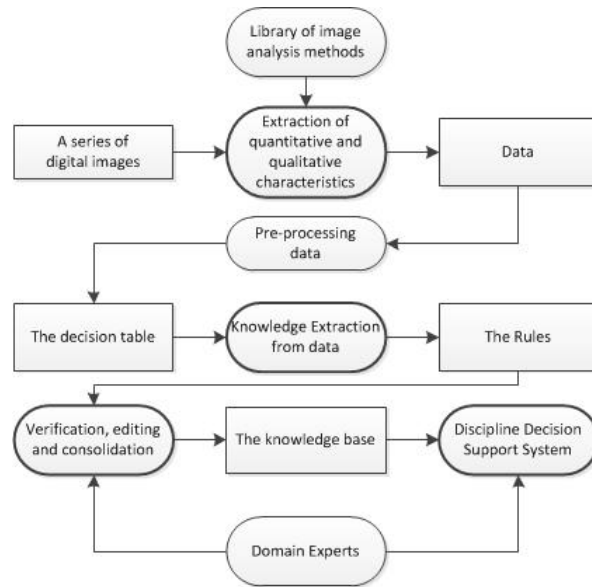


Fig. 1. Schema of two-step method of data mining from digital images

practical verification of the obtained rule base, as well as to achieve ready to implement the user system.

References

- [1] L. Shu-Hsien, C. Pei-Hui, and H. Pei-Yuan, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, pp. 11303–11311, 2012.
- [2] B. Chanda and D. D. Majumder, *Digital Image Processing and Analysis*, Second. PHI Learning Pvt. Ltd., 2011, p. 469.
- [3] I. Bankman, *Handbook of Medical Image Processing and Analysis*, Second Edition, Second. Academic Press, 2008, p. 1000.
- [4] T. D. Klaus, *Guide to Medical Image Analysis. Method and Algorithms.*, Springer L. New York, 2012, p. 487.
- [5] R. C. Gonzales and R. E. Woods, *Digital image processing*, Third. New Jersey: Prentice-Hall Inc, 2008.
- [6] Z. Pawlak, *Information System - theoretical foundation*. Warszawa: WNT, 1983.
- [7] Z. Pawlak and A. Skowron, "A rough set approach for decision rules generation," in *ICS Research Report 23/93*, p. 1993.