Image Classification Method Based on Image Informational Characteristics[1]

I. B. Gurevich and I. V. Koryabkina

Scientific Council on Cybernetics, Russian Academy of Sciences, ul. Vavilova 40, Moscow, GSP-1, 119991 Russia
e-mail: igourevi@ccas.ru; irina@motor.ru

Abstract—A method for selecting an efficient image recognition algorithm according to image informational nature is considered. An example of its application for choosing the algorithm for detecting hematological disorders and diseases, including cancers (leukemia and lymphomas), is shown.

INTRODUCTION

Images, being objects of rather specific properties, require specific analysis, processing, and recognition techniques. Traditional methods treat an image as a matrix of pixels with different brightness; they do not consider syntactic and semantic information contained in the image. In order to maximally use such information, image analysis and recognition should be based on algorithms that depend on syntactic and semantic image properties. Such a method provides perspectives for selecting image analysis algorithms and automation (partial or complete) of image processing.

This paper provides a method for selecting an efficient image recognition algorithm according to image informational nature. The fundamental hypotheses for this method are indicated. An example of the application of the method to the task of the detection of hematological disorders and diseases, including cancers (leukemia and lymphomas), is shown.

1. METHOD FOR IMAGE TRANSFORMATION SELECTION ACCORDING TO ITS INFORMATIONAL NATURE

It is suggested to divide all images into sets of similar images in order to select an optimal image processing algorithm. By introducing a similarity measure on the basis of image equivalence, one obtains the classes of image equivalence.

The proposed principles of selecting image transformation in image recognition tasks are based on a descriptive approach to image analysis [1]. The following hypotheses are used: (a) any image has a certain regularity or a mixture of regularities of different types; (b) each class of image equivalence has a corresponding class of transformations that, when applied, do not move the result of transformations out of the corresponding equivalence class.

A formalized concept of image equivalence provides a tool for decomposing an image set into subsets of images of a certain type and puts into correspondence some subsets of operations. The search for the operations in the subsets of a lower cardinality results in the (partial) elimination of an exhaustive search during the optimization of complex procedures (which are combinations of basic procedures) for finding the required image transformation [2].

The principles of optimal selection of image transformation according to the image informational nature are based, on the one hand, on the concept of redundancy of information contained in images and image equivalence classes and, on the other hand, on the hypotheses outlined above [3]. The following five basic stages are distinguished during selection of image transformations depending on the informational content of the image.

(1) Image characterization.

At this stage, the hypothesis is made about the equivalence class that the input image belongs to. Note that the classes of equivalent images strongly depend on the recognition task at hand, particularly, on transformations which supposedly do not move an image out of its equivalence class (e.g., parallel translation in character recognition) and on images that are supposed to be similar (e.g., all handwritten characters "a"), etc.

(2) Image model construction.

At the second stage, an image is reduced to a recognizable form with the help of basic operations from the corresponding class of transformations. It should be noted that there are no additional constraints such as completeness, closeness, etc., on the transformations that correspond to the class of image equivalence. It is only required that none of the algorithms that corre-

[1] This work was partially supported by the Russian Foundation for Basic Research, project no. 01-07-90016; by the Federal Target Oriented Program "Research and Development in Priority Directions of Science and Technology" in 2002–2006, project no. 37.011.11.0015; and by the Complex Program of Scientific Research of the Russian Academy of Sciences "Mathematical Modeling, Intellectual Systems, and Control of Nonlinear Mechanical Systems," project no. 2.14.

Received October 29, 2002

sponds to the class of image equivalence should move the result of recognition out of this class.

The most commonly used image models are the following: template matching, feature models, Fourier models, and structural image descriptions.

While constructing an image model, information that an image conveys can be partly lost. This makes perfect reconstruction of the input image from its model impossible, though the class of image equivalence can be determined. Since information contained in images is redundant and there is no need for all information to solve classification problems, image models are often used in applied recognition problems.

(3) Determination of the equivalence class of the image model.

Here, the image model is classified as belonging to one of the possible classes of image model equivalence. There are different ways of determining the equivalence classes of image models; e.g., by decision rules, the image feature model can be classified without additional computational cost.

The equivalence classes of the initial image and of its model are matched. If they do not coincide, the hypothesis is rejected and a new hypothesis about the equivalence class the input image belongs to is put forward. If they do coincide, then those algorithms that solve the recognition problem for this class of image models in the best way are chosen.

(4) Image model classification.

Here, the optimal image recognition algorithm is chosen among recognition operators from the class of transformations that corresponds to the equivalence class of image models. Then the problem is solved with the help of this algorithm.

Many recognition algorithms, being parametric, can be adjusted to the initial data (e.g., algorithms of estimate calculation, method of potential functions, etc.).

(5) Checking whether the image characterization was correct.

At the last stage, the equivalence class of the initial image is reconstructed using the result of recognition and inverse transformations that correspond to the equivalence classes of the images that were used. Inverse transformations can be determined differently; e.g., a typical image from the equivalence class found by the recognition algorithm for the input image is compared with the input image itself, or several images that characterize the equivalence class in the best way are selected, some weights are defined for each image, and then the initial image is compared with the weighted sum of selected images.

If the image was misclassified, a new hypothesis is made about the equivalence class the image belongs to. Then the problem is solved according to the new hypothesis. The procedure stops as soon as the problem is efficiently solved.

The practical application of the method should be divided into two stages: a learning stage, when parametric algorithms are adjusted on the training sample, and a recognition stage, when new input images are classified by using the results of the learning stage.

2. APPLICATION OF THE PROPOSED METHOD: AN EXAMPLE

In this section, the example of application of the proposed method is given for the task of detecting hematological disorders and diseases, including cancers (leukemia and lymphomas). A base of photomicrographic images of lymphatic tissue imprints was compiled to support selection and description of diagnostically important features of lymphocyte nuclei images. The base contains 1255 photographic images of specimens of 32 patients, including 24 cases of aggressive lymphoid tumors (lymphosarcoma) and eight cases of indolent chronic lymphocytic leukemia. Recognition task is defined as image classification of an input cell: whether the cell in the image is malignant or not.

As was mentioned above, the application of the proposed method was divided into two stages: learning and recognition.

2.1. Learning Stage

The analysis shows that simultaneous application of a large number of parameters allows us to objectively characterize a population of tumor cells; therefore, input cell images can be effectively described with the help of feature models. The features that optimally characterize images of each class are chosen from the set of all possible features (geometrical, logical, statistical, granulometric, and spectral) at the learning stage. Features having a high correlation ratio are discarded. Thus, by flexibly adjusting features that describe objects of each class, the maximal preservation of the essential information contained in cell images of each class is attained.

The system of decision rules is designed; it allows us to classify the obtained feature model of an input image as malignant or nonmalignant.

For the successful application of the method, parameters of the proposed algorithm are independently adjusted on the available training sample for cells from the malignant and nonmalignant classes.

2.2. Recognition Stage

In this section, a selection of image recognition algorithm depending on image informational nature is illustrated. It involves the following steps:

(1) Image characterization. The hypothesis about the class of equivalence of input image is made, whether it is an image of the nucleus of a malignant cell or not.

(2) Image model construction. At the learning stage, the optimal features for each class of cells are selected. This knowledge is used to construct a feature model of the input cell image that in the best way reflects the information contained in it. This information naturally depends on the equivalence class of the input cell image.

(3) Determination of the equivalence class the image model belongs to. By the decision rules, the image model is classified as belonging either to the class of malignant cell images or to the class of nonmalignant cell images.

(4) Image model classification with the help of recognition operators from the set of transformations that corresponds to the equivalence class of image models.

(5) Reconstruction of image equivalence class by the use of the recognition result and inverse transformations that correspond to the used equivalence class of images.

CONCLUSIONS

A method for selecting efficient image recognition algorithm according to image informational nature was elaborated. The method is based on the notion of equivalence and on the descriptive approach hypothesis that each image equivalence class can be put into correspondence with some class of transformations that, when applied to an image or its model, do not bring them out of their equivalence class. The proposed method allows for specific information content of images in choosing algorithm for image processing, analysis, and recognition. Thus, partly automated image analysis is attained.

Application of the method was demonstrated in the task of detection of hematological disorders and diseases including cancers (leukemia and lymphomas).

The following directions of future research can be mentioned:

—Exploration of different ways of defining equivalence relation on the image set.

—Investigation of correspondence between basic algorithms of image processing, analysis, recognition, and understanding and the classes of image equivalence.

—Checking a set of basic algorithms put into correspondence to some equivalence class.

—Computer experiments on selecting transformations with different methods for setting equivalence.

REFERENCES

1. Gurevitch, I.B., The Descriptive Framework for an Image Recognition Problem, *Proc. 6th Scandinavian Conf. on Image Analysis,* Oulu, 1989, in 2 vols., vol. 1. Pattern Recognition Society of Finland, 1989, pp. 220–227.
2. Gurevich, I.B. and Smetanin, Yu.G., On The Equivalence of Images in Pattern Recognition Problems, *Proc. 12th Scandinavian Conf. on Image Analysis,* Bergen, 2001, Norwegian Society for Image Processing and Pattern Recognition, 2001, pp. 679–685.
3. Koryabkina, I.V., Informational Specificity of an Image in Pattern Recognition Environment, *Proc. of the IASTED Int. Conf. in Cooperation with the Russian Academy of Sciences (Siberian Branch) on Automation, Control and Information Technology,* Novosibirsk, 2002, pp. 435–438.