

Feature Extraction Model in Systems of Diagnostics of Plant Diseases by the Leaf Images

N. Mirzaev¹, E. Saliev²

Scientific-Innovation Centre of Information-Communication Technologies at the
Tashkent University of Information Technologies Named after Muhammad al-Khorezmi,
Tashkent, Republic of Uzbekistan

E-mail: ¹nomazmirza@rambler.ru, ²salieva@gmail.com

Received: November 16, 2017

Abstract. The task of features extraction occurring in creation of systems to diagnose crop diseases using leaves images has been examined in this work. The model of diagnostic feature extraction of leaves images has been proposed for the problem solution. The main concept of the proposed model includes creation of preferred feature set simplifying the construction process of considered rule in pattern recognition given in the form of images. Experimental investigations at problem solving on diagnosis of wheat diseases with the help of leaves images have been conducted for function test of the presented model.

Keywords: diagnostic systems, base fragments of images, diagnostics of plant diseases, diagnostic features of leaves, preferred features, correlation of features.

INTRODUCTION

The issues of creation and use of information systems and technology in agricultural production under modern conditions become a key factor of innovation activities and scientific and technological progress in the field [1]. One of the major tasks in the sphere of modern IT application has to do with the creation of computer systems for identification of crop diseases and forecasting of their development.

It is known that measurement technique is a key factor of scientific and technology development in all sectors of the national economy. Nowadays, information-measuring systems (IMS) are broadly used among the various types of measuring tools due to the requirements of production, trade and other spheres of human activity. Moreover, the scope of IMS use grows continuously.

Over the last years the number of published works have been increased in the field of information systems creation, especially IMS almost in every direction of the agricultural science and production [1-14]. This fact is connected with information and analysis support of experts' intellectual activity that provides productivity in the subject area of industrial technology.

Today qualifying operations have become an integral part of manufacturing processes. The issues of creation and use of IMS in agricultural production under the modern conditions become a key factor of innovation activities in the field.

One of the main tasks in the scope of application of modern IMS has to do with the creation of information systems for identification of crop diseases, forecasting of their development and spread of crop pests. It is related to the fact that the problem of plant protection from harmful organisms is the most significant which involves the interests of the State [14]. In addition to this, it is evident that without unbiased information on pest condition, diseases and crop weeds, on the one hand, and environment and its trend change on the other hand, practical implementation of protective measures is connected with high expenditures. The application of information system of diagnostics and control phytosanitary condition allows improving reliability of information with the help of diagnostics that gives an opportunity for implementation of earlier and more accurate crop disease diagnosis and make decisions on taking measures to protect them [7–9]. In connection to this, creation of information systems of blights diagnostics (including wheat) is a vital task in controlling of crop harvest. This fact lays the groundwork for frequent occurrence of scientific researches where the issues on blight diagnosis are considered [3, 6–9, 12]. For example, the problems of crop diseases diagnostics have been examined in this work [12] using artificial neural networks. The fundamental idea of the study is a separation of color and textural features of the leaves images. The problems of information system creation of hips diseases of crop have been studied in the work [11] on the basis of image processing of their leaves. The issues of leaves images processing have been analyzed in the paper [13].

The source literature analysis, in particular [2–14], shows that the problem of crop disease diagnosis has not been enough studied. Therefore, the objectives associated with the questions of information extraction on the examined plants given in the form of leaves images are still important today.

The distinctive feature of the studied approach to the problem solution of features selection is the set formation of preferential characteristics simplifying the construction of decision procedure in the object recognition set in the form of pictures. Moreover, it helps to accomplish transition from specific algorithms of feature selection to the model – a family of algorithms for uniform description of the decision procedure. It should be noted that the present paperwork represents an updated and revised version [14].

The aim of the study is the development of formation model of diagnostic features defining crop disease by source picture of leaves. The approach which is used here is based on the statistical analysis of leaves pictures of the diagnosed crops.

In order to gain the above set goal it is necessary to solve the following tasks:

- 1) to make reappraisal of the existed approaches to problem solving of crop diseases by the leaves photos and define the research objectives;
- 2) to develop a model of the image feature extraction associated with identification of crop blights by the leaves pictures;
- 3) to carry out experimental research for performance evaluation of the developed model.

From a scientific perspective the findings of the work represent a new solution to the scientific learning connected with the object recognition submitted in a form of images. The practical significance of the obtained results is that the engineered algorithms and programs can be used in solving of applications in relation to the object identification presented in the form of images (for instance, in the problems of crop diseases diagnosis with the help of leaves pictures, person identification by a portrait).

STATEMENT OF THE PROBLEM

We consider the set of admissible objects \mathfrak{R} , given as images of plants leaves. Initial data on each admissible object (investigated plant) \mathfrak{S} are given three dimensional matrix

(color image) X with the size $c \times m \times n$ (where c – the number of color channels; m and n – the numbers of rows and columns respectively):

$$X = \begin{pmatrix} x_{c11} & \dots & x_{ci1} & \dots & x_{cn1} \\ \dots & \dots & \dots & \dots & \dots \\ x_{c1j} & \dots & x_{cij} & \dots & x_{cnj} \\ \dots & \dots & \dots & \dots & \dots \\ x_{c1m} & \dots & x_{cim} & \dots & x_{cnm} \end{pmatrix}. \quad (1)$$

It is assumed that the objects of the set \mathfrak{R} are divided into two disjoint subsets (of the class) K_1 and K_2 [15]:

$$\mathfrak{R} = \bigcup_{j=1}^2 K_j, \quad K_1 \cap K_2 = \emptyset. \quad (2)$$

The division \mathfrak{R} is not completely defined. There is only some initial information J_0 about the classes K_1, K_2 . Let there be some sample $\tilde{\mathfrak{S}}^m$ ($\tilde{\mathfrak{S}}^m \subset \mathfrak{R}$) consisting of m objects ($\tilde{\mathfrak{S}}^m = \{\mathfrak{S}_1, \dots, \mathfrak{S}_i, \dots, \mathfrak{S}_m\}$, $\mathfrak{S}_i \in \mathfrak{R}$, $i = \overline{1, m}$):

$$\tilde{K}_j = \tilde{\mathfrak{S}}^m \cap K_j, \quad C\tilde{K}_j = \tilde{\mathfrak{S}}^m \setminus \tilde{K}_j. \quad (3)$$

Then the initial information J_0 about the classes can be given in the form [15]:

$$J_0 = \{\mathfrak{S}_1, \tilde{\alpha}(\mathfrak{S}_1); \dots; \mathfrak{S}_i, \tilde{\alpha}(\mathfrak{S}_i); \dots; \mathfrak{S}_m, \tilde{\alpha}(\mathfrak{S}_m)\}, \quad \tilde{\alpha}(\mathfrak{S}_i) = (\alpha_{i1}, \alpha_{i2}), \quad (4)$$

where α_{ij} – value of the predicate $P_j(\mathfrak{S}_i) = "\mathfrak{S}_i \in K_j"$.

The task is to build such an operator that allows to form a space of features that characterizes the diseases of cultivated plants according to the initial image of the leaves. The determination of the set of diagnostic features $\{\tau_{u1}, \dots, \tau_{ui}, \dots, \tau_{um}\}$ is carried out according to the given initial information J_0 on the basis of statistical analysis. In this case, it is required that the formed feature space be smaller than the original one, and it should ensure the separation of the specified objects with some accuracy.

PROPOSED APPROACH

A non-standard approach to the problem of feature set formation identifying crop diseases has been studied in the research. The model of feature extraction of the object set in the form of images has been offered on the basis of this approach. A fundamental concept of the proposed model involves space creation of independent (or weakly dependent) diagnostic features in pattern recognition given in the form of images. Therewith three-dimensional (two-dimensional) structure of leaves images is introduced as one-dimensional space of characteristics of high dimensionality. Each feature defines only a specific area (fragment) of a picture. Various statistical characteristics are used for description of each fragment of an initial picture.

The proposed model of operators for extracting the diagnostic features of images of leaves of cultivated plants includes the following main steps.

1. *Formation of basic fragments of images of leaves.* The first step in setting the model for the extracting of features is the formation of a system of basic fragments of images Σ , where $\Sigma = \{\Xi_1, \dots, \Xi_k\}$, depending on the parameter k , where $k = m_H \times n_W$, m_H is the number

of vertical divisions of the image, and n_w is the number of image divisions by width. This parameter indicates the fixed power of the subsets that are formed as a result of the division of the image in question. By specifying different integer values for this parameter, different amounts of base fragments can be obtained. This requires that the number of elements (pixels) in each image fragment should be the same. As a result of this stage, k the same rectangular fragments are formed. If the dimension of the original image is $H \times W$ pixels, then the resulting image fragments will consist of $k_H \times k_W$ pixels:

$$\begin{aligned} k_H &= H \operatorname{div} m_H + \delta_H, & k_W &= W \operatorname{div} n_W + \delta_W, \\ \delta_H &= \begin{cases} 0, & \text{if } H \bmod m_H = 0, \\ 1, & \text{else;} \end{cases} & \delta_W &= \begin{cases} 0, & \text{if } W \bmod n_W = 0, \\ 1, & \text{else.} \end{cases} \end{aligned} \quad (5)$$

If there are residuals from division, the sizes of image fragments are increased by one pixel, the beginning of the others is shifted to the left (up) by one pixel.

Depending on the method of forming the system of basic fragments Ξ_q , where $q = \overline{1, k}$, we can get a variety of feature extraction algorithms [16].

2. *Determining of a set of diagnostic features of leaves.* At this stage, a set of diagnostic features is formed, which are defined as sampling points on the reference subset in question. It is known [17] that the moment of order n_p , where $n_p = (p + q)$, of the random variable $f(x, y)$, where $(x, y) \in \Xi_u$, is defined as

$$m_{pq} = \frac{1}{N_u} \sum_{(x,y) \in \Xi_u} x^p y^q f(x, y), \quad N_u = |\Xi_u|, \quad p, q = 0, 1, \dots, n_p, \quad (6)$$

here n_p is a parameter indicating the maximum order of the computed moments.

To determine the central moments, we use the following formula:

$$\mu_{pq} = \frac{1}{N_u} \sum_{(x,y) \in \Xi_u} (x - \bar{x}_u)^p (y - \bar{y}_u)^q f(x, y), \quad (7)$$

where \bar{x}_u and \bar{y}_u coordinates of the central point Ξ_u .

In addition to the characteristics considered, it is possible to calculate variance, autocorrelation, etc., as diagnostic features of the object.

As a result of this stage, we get a set of diagnostic features. The generated feature space will be denoted by Y ($Y = (y_1, \dots, y_{k \times n_p})$).

3. *Extracting of subsets of strongly correlated features.* At this stage, a system of “independent” subsets of diagnostic features Y is defined.

The subsets of strongly correlated features are extracted as follows. Let $\Gamma_u (u = \overline{1, k'})$ - be subsets of diagnostic strongly correlated features. The proximity measure $L(\Gamma_u, \Gamma_v)$ between the subsets Γ_u and Γ_v may be given with different ways, for example:

$$L(\Gamma_u, \Gamma_v) = \frac{1}{N_u \cdot N_v} \sum_{y_i \in \Gamma_u} \sum_{y_j \in \Gamma_v} \eta(y_i, y_j), \quad (8)$$

where N_u, N_v - the number of diagnostic features entering respectively into subsets Γ_u and Γ_v ; $\eta(y_i, y_j)$ - function characterizing the strength of the pair relationship between the signs y_i and y_j .

As a result of this stage, a set of "independent" subsets of strongly correlated features $W_A = \{\Xi_1, \Xi_2, \dots, \Xi_n\}$ is determined. The number of extracted subsets will depend on the parameter n . By setting this parameter to various integer values, we can get different algorithms [18].

4. *Determining of preferred features.* We consider the subsets $\{\Xi_1, \dots, \Xi_q, \dots, \Xi_n\}$, which are defined in the previous step. Let N_q be the power of a subset of strongly correlated features. A selection from a subset Ξ_q the preferred feature is made on the basis of an evaluation of the dominance of the feature in question, which divides the objects belonging to the set $\tilde{\mathfrak{S}}^m$ into two subsets (classes) K_1 and K_2 [19]:

$$D_i = \frac{\hat{N}_2 \sum_{j=1}^2 \sum_{\mathfrak{S} \in \tilde{K}_j} \sum_{\mathfrak{S}_u \in \tilde{K}_j} (a_i - a_{iu})^2}{\hat{N}_1 \sum_{\mathfrak{S} \in \tilde{K}_1} \sum_{\mathfrak{S}_u \in \tilde{K}_2} (a_i - a_{iu})^2}, \quad (9)$$

where $\hat{N}_1 = (m_1(m_1 - 1) + m_2(m_2 - 1))/2$, $\hat{N}_2 = m_1 \times m_2$, $m_1 = |\tilde{K}_1|$, $m_2 = |\tilde{K}_2|$.

The smaller the value D_i , the greater the preference is given to the corresponding feature. If two or more features receive the same preference, then any one of them is selected. In calculating D_i , it is assumed that \mathfrak{S} and \mathfrak{S}_u are different objects (i.e. $\mathfrak{S} \neq \mathfrak{S}_u$).

For each subset of strongly correlated features Ξ_q , a preferred feature is identified at this stage, which is denoted by χ_q . As a result, the following set of preferred features is formed: $\chi_1, \chi_2, \dots, \chi_n$.

Each preferred feature from this set is representative of only one subset of strongly correlated features.

Thus, a model of operators for the formation of diagnostic features for the image of leaves of cultivated plants is defined. To assess the efficiency of the model, experimental studies were conducted.

In order to assess the working capacity of the model, functional schemes and corresponding algorithms have been developed that determine the structure of the program to be created.

EXPERIMENT AND RESULTS

Functional diagrams and relevant diagnosis programs have been developed for common use and performance evaluation of the analysed model. The elaborated programs have been implemented in Delphi language. For the purposes of functional test of the developed programs we shall consider the diagnosis task of the stripe rust of wheat using leaves images.

The fact is that rust diseases of grain, especially wheat, are more harmful and dangerous in many parts of the world. The injuriousness of these wheat diseases and volumes of harvest losses depend on a number of factors, primary lesion period (i.e. phases of wheat growth, starting time of a disease), growth rate and others.

A precise determination of phase of development is of great importance not only in the analysis of harmful rust diseases but also in the conducting of research on forecasting of disease development and plan organization for plant protection.

A precise determination of phase of development is of great importance not only in the analysis of harmful rust diseases but also in the conducting of research on forecasting of disease development and plan organization for plant protection.

The collection of 300 wheat leaves images is given as initial data (i.e. original sample V). Furthermore, a number of possible diagnoses (phytosanitary state) equals 2:

- wheat leaves images with detected stripe rust (K_1);
- wheat leaves images without stripe rust (K_2).

The first subset K_1 has 150 pictures and the second – the rest. These pictures have been divided into educational and check samples. The method of cross-validation is applied for accuracy calculation at problem solving for the exception purposes of successful (or unsuccessful) partition of the original sample V into two parts V_o and V_k ($V = V_o \cup V_k$, where V_o – training set, V_k – test set) [18, 20]. In this technique the original sample of objects is split into 30 non-crossing blocks in a random manner containing 10 objects each. It is required to keep proportion in all blocks quantitatively to objects belonging to different classes. As the result 5 objects of each class can be observed in each block. The procedure of cross-validation in these blocks includes several steps. 22 from 30 blocks are chosen on each step in the function of educational sample, and the model with set parameters is taught at this sampling. The model taught in such a way is checked in other 8 blocks (check sample). The model quality rating on accuracy classification is identified and recorded at the end of each control. Each block one by one is selected and their positions are changed when implementing every next step from test and training sets. The appropriate blocks are marked and they do not participate when selecting candidates for including control sample to exclude objects reuse of check samples. The correctness of diagnosis is defined as average after finishing the procedure of cross-validation exam.

The task of diagnostics has been completed with the help of: 1) the model described in [11]; 2) the proposed model of feature extraction. Experimental results in problem solving with the use of these models are presented, correspondingly, in the Tables 1 and 2.

Table 1. The results of solving the diagnostic problem using the known model

Diagnosis	The average number of correct diagnoses	The average number of incorrect diagnoses	Average number of failures	Accuracy of diagnosis
Sick	32.1	6.4	1.5	80.3 %
Not sick	34.6	4.5	0.9	86.5 %

Table 2. The results of solving the diagnostic problem using the developed model

Diagnosis	The average number of correct diagnoses	The average number of incorrect diagnoses	Average number of failures	Accuracy of diagnosis
Sick	36.3	2.9	0.8	90.8 %
Not sick	37.6	1.9	0.5	94.0 %

According to Table 1, when using the model considered in [11], 66.7 objects from 80 objects (images of leaves) were properly identified on average, which is 83.4% recognition accuracy. Using the developed model, 73.9 objects were correctly identified from 80 objects (images of leaves), which is 92.4% recognition accuracy (Table 2).

The carried out experimental researches have shown the efficiency of the developed model of extraction of diagnostic features in solving the given task. As a result of the experiment, a set of diagnostic features is generated, which allows to split the objects of the control sample into two classes with an acceptable error.

CONCLUSION

The issues on diagnostics of crop diseases and data support of decision-making on their protection are one of the main in the problem of harvest control. Nevertheless, tasks connected with the development of automated systems of crop diseases diagnostics are not studied well.

The model of diagnostic feature creation has been developed in the detection of blights using leaves images. Moreover, formation of diagnostic features is supported by calculation of different statistical characteristics for each fragment of the initial picture.

In the course of solving a practical task it has been specified that the steps of subset formation of "independent" features, especially, issues on detection of a number of such subsets and collection of diagnostic features by leaves pictures as well as extraction of preferable characteristics play the most important role in task solving of diagnostics. Therefore, it is necessary to continue research taking into account the ascertained directions.

The developed model can be employed in composition of diverse software applications focused on decision-making of objects classification prescribed in the form of images.

REFERENCES

1. Alt, V. V. (2006). Informatsionnyye tekhnologii kak faktor povysheniya effektivnosti vybora tekhnologicheskikh resheniy [Information technologies as a factor in increasing the efficiency of the choice of technological solutions]. *Dostizheniya nauki i tekhniki APK [Achievements of Science and Technology of AIC]*, 2006(11), 3–5 (in Russian). Available at: <https://elibrary.ru/item.asp?id=10333014>.
2. Li, K., Xiong, L., Zhang, D., Liang, Z., & Xue, Y. (2017). The research of disease spots extraction based on evolutionary algorithm. *Journal of Optimization*, 2017, article ID 4093973, 1–14. doi:10.1155/2017/4093973.
3. Kadakol, S., & Maned, J. (2016). Intelligence system for leaf extraction and disease diagnostic. *International Journal of Computer Sciences & Engineering Engineering*, 4(3), 62–66. Retrieved from: http://www.ijcseonline.org/pdf_spl_paper_view.php?paper_id=64&14-I009-IJCSE.pdf.
4. Kaur, L., & Laxmi, V. (2016). A review on plant leaf classification and segmentation. *International Journal of Engineering and Computer Science*, 5(8), 17658–17661.
5. Shripriya, R., & Yuvaraj, N. (2016). A survey on leaf disease prediction algorithms using digital image processing. *Indian Journal of Innovations and Developments*, 5(10), 1–4. Retrieved from: <http://ijid.informaticspublishing.com/index.php/ijid/article/view/111065>.
6. Patil, R., Udgate, S., More, S., Nemishte, D., & Kasture, M. (2016). Grape leaf disease detection using k-means clustering algorithm. *International Research Journal of Engineering and Technology*, 3(4), 2330–2333.
7. Dange, D. J., & Sayyad, M. A. (2015). Computer vision image enhancement and plant leaves disease detection. *International Journal of Modern Trends in Engineering and Research*, 2(6), 106–110. Retrieved from: <https://www.ijmter.com/papers/volume-2/issue-6/computer-vision-image-enhancement-and-plant-leaves-disease-detection.pdf>.
8. Petrellis, N. (2015). Plant disease diagnosis based on image processing, appropriate for mobile phone implementation. In *Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015)* (pp. 238–246). Kavala, Greece. Retrieved from: http://ceur-ws.org/Vol-1498/HAICTA_2015_paper28.pdf.
9. Arivazhagan, S., Newlin Shebiah, R., Ananthi, S., & Vishnu Varthini, S. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal*, 15(1), 211–217. Retrieved from: <http://www.cigrjournal.org/index.php/Ejournal/article/view/2338/1720>.
10. Mondal, K. K., & Shanmugam, V. (2013). Advancements in the diagnosis of bacterial plant pathogens: An overview. *Biotechnology and Molecular Biology Review*, 8(1), 1–11. doi:10.5897/BMBR12.007.

11. Anami, B. S., Pujari, J. D., & Yakkundimath, R. (2011). Identification and classification of normal and affected agriculture/horticulture produce based on combined color and texture feature extraction. *International Journal of Computer Applications in Engineering Sciences*, 1(3), 356–360.
12. El-Helly, M., Rafea, A., & El-Gammal, S. (2003). An integrated image processing system for leaf disease detection and diagnosis. In *Proceedings of the 1st Indian International Conference on Artificial Intelligence (IICAI 2003)* (pp. 1182–1195). Hyderabad, India.
13. Denisyuk, V. S. (2008). Algoritmy vydeleniya osobennostey na izobrazheniyakh s tsel'yu klassifikatsii zabolvaniy rasteniy [Algorithms of image analysis for feature detection problem in order to diagnose plants diseases] (in Russian). In V. N. Kasyanov (Ed.), *Konstruirovaniye i optimizatsiya parallel'nykh programm [Parallel programs construction and optimization]* (pp. 71–82). Novosibirsk, Russia: A. P. Ershov Institute of Informatics Systems. Retrieved from: https://www.iis.nsk.su/files/articles/sbor_kas_16.pdf.
14. Mirzaev, N. M. (2012). Model' vydeleniya priznakov v zadache diagnostiki fitosostoyaniya rasteniy po izobrazheniyam list'yev [Feature extraction in the problem of diagnosing of plants' phytocondition by leaves' images] (in Russian). *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta [Vestnik of Ryazan State Radio Engineering University]*, 3(41), 21–25.
15. Zhuravlev, Yu. I. (1998). *Izbrannyye nauchnyye trudy [Selected papers]* (in Russian). Moscow, Russia: Magistr.
16. Mirzaev, N. M. (2010). About one model of image recognition. *Computer Technology and Applications: Proceedings of The First Russia and Pacific Conference* (pp. 394–398). Vladivostok, Russia.
17. Gonzales, R., & Woods, R. (2008). *Digital image processing*. New Jersey, USA: Prentice Hall.
18. Fazilov, Sh. Kh., Mizraev, N. M., & Mizraev, O. N. (2016). Postroyeniye raspoznyayushchikh operatorov v usloviyakh vzaimosvyazannosti priznakov [Building of recognition operators in condition of features' correlations] (in Russian). *Radioelektronika, informatika, upravleniye [Radio Electronics, Computer Science, Control]*, 1(36), 58–63. Retrieved from: http://nbuv.gov.ua/UJRN/riu_2016_1_9.
19. Kamilov, M. M., Fazilov, Sh. Kh., Mirzaev, N. M., & Radjabov, S. S. (2012). Estimates calculations algorithms in condition of huge dimensions of feartures' space. *Problems of Cybernetics and Informatics (PCI'2012): Proceedings of the 4th International Conference, vol. I* (pp. 184–187). Baku, Azerbaijan.
20. Braga-Neto, U. M., & Dougherty, E. R. (2016). *Error estimation for pattern recognition*. New York, USA: Springer.