UNITECH 2015 GABROVO

INTERNATIONAL SCIENTIFIC CONFERENCE
20 – 21 November 2015, GABROVO

CLASSIFICATION OF RICE GRAINS USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

Ozan AKI
Trakya University

Aydın Güllü
Trakya University

Erdem Uçar
Trakya University

Abstract

Productive agriculture needs technologies. Technologies are using in agriculture in several stages. In this study, classification of rice grains were tested using image processing and machine learning techniques. There are three grain types used for classification. But broken grains were considered as a type too. Thus total of 4 grain types were used for classification. Rice grain images obtained by a webcam. Six attribute were extracted for each grain image. Grain attributes are related to its shape geometry. Attributes of known types of rice grains were used for training several machine learning algorithms. Most success nearest neighbor with generalization algorithm was selected for real time testing. Real time test classification result's accuracy were calculated as 90.5% with selected classifier.

Keywords: classification, rice, grain, image processing, machine learning.

INTRODUCTION

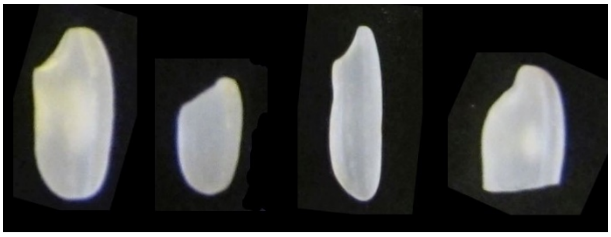
Technologies using in agriculture are very important for more productive and sustainable production. Some of the usage areas of these technologies are quality control and classification of grains. Classifying the similar small grains can be possible by using image processing. Also processing a lot of grains concurrently can be possible by camera viewing angle [1, 2].

In this study, we aim to classify rice grains by using image processing and machine learning techniques. Rice grains images were obtained by a camera. Each rice grain in image were segmented separately. Six attribute were extracted from each grain image related to its shape geometry [1]. An ARFF (Attribute Related File Format) file was built by using these values of known type of rice grains [3].

ARFF file was used for evaluate the some machine learning algorithms. After evaluating machine learning algorithms, most success one is selected for real time testing. Real time test were made by predicting class of unknown type of rice grains. Classified rice grains in image were counted and colored by difference of its class. Then results were compared with real situation.

MATERIAL AND METHODS

Three type of rice grains were used for classification. Also, each type's broken grains were considered as a type separately. Therefore, number of 4 types of rice grain used for classification in this study. Used finished rice grains were bought from a local market. Used rice grain type names are *baldo*, *osmancık*, *yasemin* named in Turkish. Also, broken grains were classified as *broken* type. Grain samples were shown on Figure 1.



(a) (b) (c) (d)

Fig. 1. *Types of rice grains a) Baldo b) Osmancik c) Yasemin d) Broken*

Grains were spread out on solid black platform and their images were obtained by a webcam. Intersection of grains with each other were prevented. Each image frame taken from webcam has about 80 to 100 grains related to its type.

Obtained images were processed in several stages. First stage of processing image is applying filters for smooth image and reduce noise. In this study, firstly grabbed image converted to gray scale image and Gaussian filter with 3x3 kernel size and pyramid down and pyramid up filters applied to grabbed image. Then binary image was produced by threshold image with proper value. Optionally, erode and dilate operations were applied for break the touching grain images [4, 5].

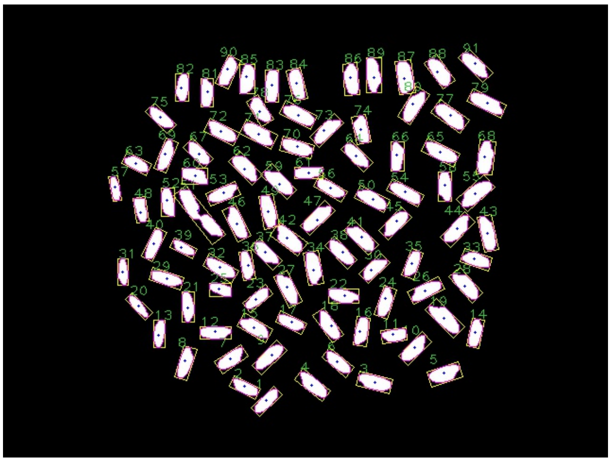


Fig. 2. *Processed image of the rice grains*

After preparing image for processing, each grain image were extracted from image by using its contour edges. Grain boundary is represented by collection of connected points. This closed polygon were used for extraction of attributes of grain shape. Extracted attributes are perimeter of grain, area, length, thickness and lack of shape from fitted rectangle. These six attributes were calculated for each rice grain in the single image [4, 5].

Calculated attributes were saved in an ARFF file for each grain with its class name. ARFF file is using by Weka machine learning application. Weka is an application that have collection of data mining tasks [3].

In this study, Weka application is used for evaluate some machine learning algorithms with collected grain attributes. Evaluation of algorithms were done with its default parameters defined in Weka application.

Developed application were used for collecting samples, extracting attributes, creating ARFF file, training machine learning algorithm and predicting real time images.

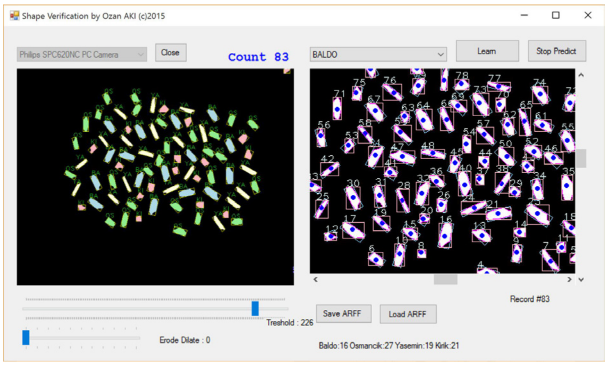


Fig. 3. *Developed application's interface*

RESULTS AND CONCLUSION

NNge (Nearest Neighbor with Generalization), NBTree (Decision Tree with Naive Bayes), RBFNetwork (Normalized Gaussian Radial Basis Function Network), KStar (Instance-based classifier), BFTree (Best-First Decision Tree), Bagging, Random Forest, J48, IB1 (Nearest-Neighbour classifier), IBk (K-Nearest Neighbours classifier), JRip (Propositional Rule Learner, Repeated Incremental Pruning to Produce Error Reduction) and Naïve Bayes machine learning algorithms were evaluated with collected gran attributes [3, 6]. Weka application were used for evaluation of machine learning algorithms. Evaluation results were shown in Table 1.

Table 1. *Evaluation results of machine learning algorithms*

Classifier	Accuracy (%)	Kappa Statistic	F-Measure
NNge	91.03	0.88	0.91
NBTree	91.03	0.88	0.91
RBFNetwork	91.03	0.88	0.91
KStar	90.81	0.88	0.91
BFTree	90.81	0.88	0.91
Bagging	90.81	0.88	0.91
Random Forest	90.81	0.88	0.91
J48	90.14	0.87	0.90
IB1	90.14	0.87	0.90
IBk	90.14	0.87	0.90
JRip	90.14	0.87	0.90
Naïve Bayes	89.46	0.86	0.89

Accuracy is defined as measuring the percentage of correct predictions with respect to the overall data. The f-measure metric combines precision and recall by calculating their harmonic mean. Recall and precision metrics indicates that errors of classifying negative instances as positive and classifying positive instances as negative respectively. Kappa has a range between -1 and 1, where -1 is total misclassification and 1 is 100% accurate classification [6, 7].

Table 1's rows were sorted by accuracy descending order. By inspecting results, values are seen as very closer to each other. NNge is one of most success algorithms and it was selected for predicting unknown types.

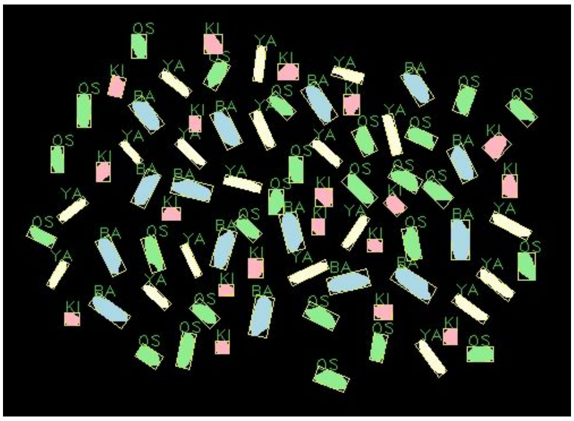


Fig. 4. *Classified rice grains indicated by colors*

NNge algorithm was trained in developed application by previously created ARFF file. Algorithm was tested in real time environment by obtaining images from webcam. Test grains were mixed from number of 20 of each type. Then predicting class of each grain image using selected algorithm. Each rice grain image was counted and colored by its defined class color. Real time test results were shown in Table 2.

Table 2. *Real time test results*

Type of Rice Grain	Real Number of Rice Grain	Predicted Number of Rice Grain	Calculated Accuracy (%)
Baldo	20	18	90.0
Osmancık	20	21	95.2
Yasemin	20	18	90.0
Broken	20	23	86.9

Average accuracy of real time test is 90.5% and this result is very close to evaluation result of used algorithm.

Note that, these learning samples and predicting results were obtained under same environment and lighting conditions. Environmental conditions especially lighting is very important in image processing and will directly affect the results. Therefore, inspected grain images must capture under well-designed conditional environment. Isolating from natural lighting and illuminating with controllable lights is first step of image processing.

REFERENCES

[1] S. Kaur and D. Singh, "Geometric Feature Extraction of Selected Rice Grains using Image Processing Techniques," *International Journal of Computer Applications,* vol. 124, 2015.

[2] R. Birla and A. P. S. Chauhan, "An Efficient Method for Quality Analysis of Rice Using Machine Vision System," *Journal of Advances in Information Technology Vol,* vol. 6, 2015.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, pp. 10-18, 2009.

[4] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*: " O'Reilly Media, Inc.", 2008.

[5] Intel. (2000). *Open Source Computer Vision Library Reference Manual*.

[6] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2005.

[7] Y. Kılıçaslan, E. S. Güner, and S. Yıldırım, "Learning-based pronoun resolution for Turkish with a comparative evaluation," *Computer Speech & Language,* vol. 23, pp. 311-331, 2009.