Methodology for determining the apple variety based on computer processing of digital images

J. Alikhanov, A. Moldazhanov, A. Kulmakhambetova, D. Zinchenko, A. Azizov, A. Nurtuleuov, D. Sarsenbekuly
Kazakh National Agrarian Research University, Almaty, Kazakhstan
E-mail: alikhanov.d@lust.ru

Abstract: *This paper presents the development and experimental validation of a method for automatic identification of apple varieties based on the analysis of visual features extracted from digital images. The proposed approach uses classical computer vision techniques without applying neural networks or deep learning, which makes the system interpretable, lightweight, and reproducible for laboratory and industrial use.*

The algorithm includes the stages of image acquisition, preprocessing, object segmentation, feature extraction, and classification using statistical models such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and logistic regression. The extracted features include geometric parameters (area, perimeter, circularity, eccentricity, axis ratio) and color characteristics (mean HSV values, red color percentage, hue distribution).

Experimental validation was performed on a dataset containing five apple varieties: Sinap Almaty, Fuji, Brebourne, Gold Delicious, and Hybrid. The system achieved an average classification accuracy of 90%, with the highest results for varieties with distinctive morphological or color characteristics. Comparative analysis with manual sorting demonstrated significant advantages in terms of processing speed, objectivity, and scalability.

The proposed method can be implemented on compact single-board computers, making it suitable for mobile quality control stations and automated sorting lines. Future work includes the integration of weight and texture parameters and the expansion of the variety database for broader applicability.

KEYWORDS: *APPLE, VARIETY, COMPUTER VISION, ALGORITHM, PROGRAM, SIZE, COLOR.*

1. Introduction

Apples are one of the main fruit crops in Kazakhstan and many other countries with a developed agricultural sector. Their variety has a decisive influence on the consumer properties, storage, transportation and price of the product. In this regard, the task of prompt and reliable varietal identification of apples is of particular importance in the context of the digitalization of agriculture and the introduction of Smart Farm technologies.

According to the Ministry of Agriculture of the Republic of Kazakhstan, there are more than 1,600 farms engaged in industrial horticulture in the country, the total area of apple plantations is 35.1 thousand hectares, of which 80% are in the Almaty and Turkestan regions. Despite the positive dynamics — an increase in gross harvest by 46 thousand tons over the past three years — the country's self—sufficiency in apples is about 70.5%, and consumption is about 12 kg per person per year, which is lower than in the leading countries (Poland — 67 kg, Turkey - 35 kg) [1-3].

Against this background, the task of automating fruit sorting and classification is becoming more urgent, especially in conditions of mass production, where manual labor becomes ineffective due to the subjectivity of the assessment., it takes a lot of time and has limited repeatability of results. The most promising direction is the application of computer image processing, which makes it possible to obtain quantitative characteristics of fruits in color, shape and texture.

A number of studies, including [4-7], demonstrate that technical vision methods provide high accuracy in measuring geometric parameters (width, height, projection area), color analysis (in fractions of the dominant color or in the HSV color space), as well as detecting visual defects. At the same time, technologies such as:

• image processing using OpenCV and Python are most often used.;
• extraction of shape features: roundness, eccentricity, compactness;
• Color analysis: calculation of the average color, the proportion of red/green/yellow components;
• the use of classifiers based on statistical models (SVM, k-NN, logistic regression).

Unlike the resources of expensive neural network solutions that require massive, labeled datasets and high computing power, classical algorithms offer a more accessible and interpretable alternative, especially applicable to small and medium-sized agricultural enterprises.

Thus, the purpose of this study is to develop and test a method for determining the apple variety based on visual features obtained from digital images, followed by classification using classical analysis methods. The proposed approach focuses on practical applicability, reproducibility, and integration into existing production lines without the need to implement neural network technologies.

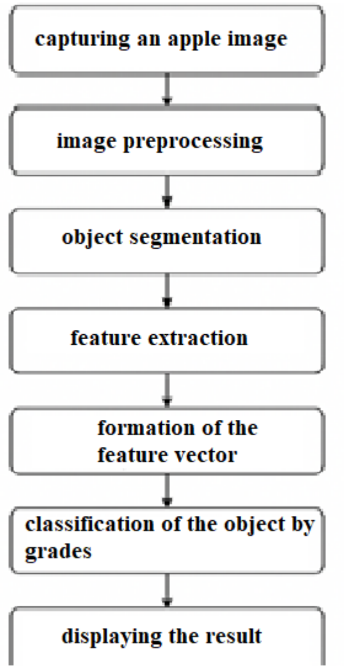

Figure 1 *— Block diagram of the algorithm*

2. Materials and methods

The developed apple automated variety identification system implements a full cycle of digital image analysis: from frame capture to classification and output of the result to the operator. The main task is to obtain and interpret the visual signs of fruits that best reflect the varietal affiliation. All processing is performed using traditional methods of technical vision without involving neural networks.

System components:

• Digital camera (12 MP);
• Controlled lighting setting (5500K daylight);
• uniform background (matte white);
• Raspberry Pi 4B or PC;
• Python 3.10 and libraries OpenCV, NumPy, Scikit-learn, PyQt5 (as required);
• interface with output of parameters and grade labeling.

The algorithm for determining the apple variety has been developed based on the methods of technical vision using digital image processing. The general flowchart of the algorithm is shown in Figure 1. It includes sequential steps from image capture to obtaining a classified result.

Algorithm Stages:

1. Capture an image of an apple using a digital camera;
2. Image preprocessing: color space transformation, binarization, filtering;
3. Object segmentation and Area of interest allocation (ROI);
4. Extraction of features: morphological and color;
5. Formation of the feature vector;
6. Classification of the object by grades based on statistical methods;
7. Display the result on the operator

The fundamental difference from neural network approaches lies in the complete interpretability of each stage and the absence of the need for deep learning on large datasets.

To implement the algorithm, a compact sorting unit was used, which includes the following components:

• a digital camera with a resolution of 12 megapixels mounted on an adjustable tripod;
• Uniform white background with non-slip surface;
• 5500K single illumination, providing stable color reproduction;
• a personal computer with Ubuntu/Python OS installed or a single-board computer (for example, Raspberry Pi 4B);
• the management interface implemented on PyQt5 (in the basic version — text output to the terminal).

The program code is implemented in Python 3.10, using the following libraries:

• OpenCV — for image processing;
• NumPy — for working with feature vectors;
• Scikit-learn — for classification;
• Matplotlib — for data visualization.

Image capture and preparation

The shooting was carried out in laboratory conditions. Each fruit was photographed from three angles: frontal, lateral and upper. Samples of varieties: Sinap Almaty, Fuji, Brebourne, Gold Delicious, Hybrid.

Image Preprocessing

Preprocessing is aimed at eliminating the background and improving the quality of color information.

Stages:

• Conversion to HSV color space
• Gaussian blur to eliminate noise
• Color filtering (by hue, saturation)
• Morphological operations (closing, opening)
• Contour analysis

Contour segmentation and analysis

Segmentation is performed by searching for the largest external contour. The contour is used to calculate the area, perimeter, moments, eccentricity, and other features.

Extracting color features

The average values of the HSV and RGB components inside the fetal mask are extracted, as well as saturation statistics and the proportion of red color.

Formation of the feature vector

The feature vector combines morphological and color indicators.

Classification of varieties

For classification, it is used:

• Support vector machine (SVM, RBF core);
• K-nearest neighbors algorithm (optimal k = 5);
• Logistic regression as a basic classifier.

Metrics and quality assessment

The accuracy is assessed using the following metrics:

• Accuracy
• Precision
• Recall
• F1-score

Visualization is the result of color segmentation with background removal;

Segmented fruit is a highlighted outline of an apple for feature extraction.

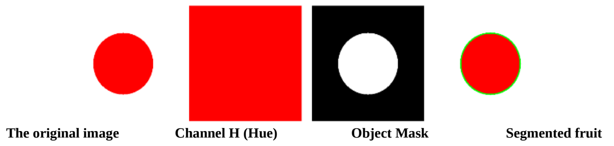

***Figure 2** – Apple image processing method*

The original image is the subject of the survey (in this example: a synthetic "apple");

Channel H (Hue) — determining the shade of the fruit for color filtering;

The object mask is the result of color segmentation with background removal.;

Segmented fruit is a highlighted outline of an apple for feature extraction.

3. Results and discussion

As a result of the experiments, digital images of apples of five varieties were processed, with the extraction of visual features and subsequent classification. The methodology was tested on a sample of 100 images (20 images per grade) that did not participate in the training.

The classification accuracy was assessed using standard methods: Accuracy, Precision, Recall, F1-score, as well as using the error matrix (confusion matrix) shown in Figure 1.

Grade (according to the expert)	**Defined correctly**	**Mistakes**	**Accuracy (%)**
Brebourne	16	4	80.0
Hybrid	17	3	85.0
Gold Delicious	20	0	100.0
Fuji	18	2	90.0
Sinap Almaty	19	1	95.0
Average accuracy	–	–	**90.0**

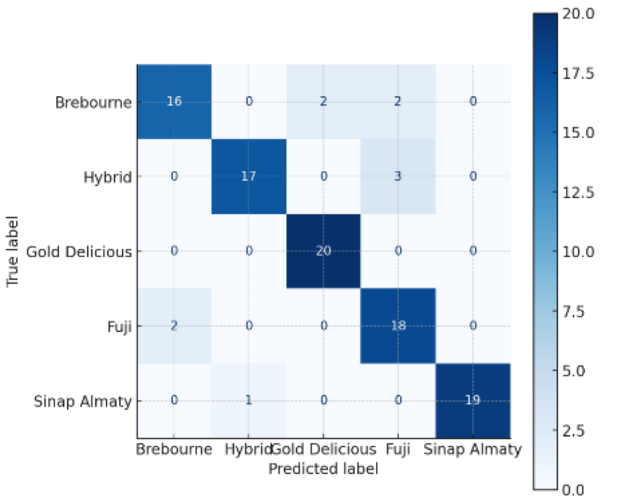

***Figure 3** – The matrix of apple classification errors by variety showed high recognition accuracy, especially for clearly defined varieties (Gold Delicious, Sinap Almaty).*

The main errors occurred when cross-classifying varieties similar in morphology, such as Brebourne and Fuji. This is due to the similar values of geometric features with minor color variations.

Morphological analysis allowed us to quantify the differences between the varieties. According to the results of the analysis, roundness and eccentricity are reliable signs of differentiation. Gold Delicious had the highest roundness values, which is confirmed by visual measurements (Fig. 2).

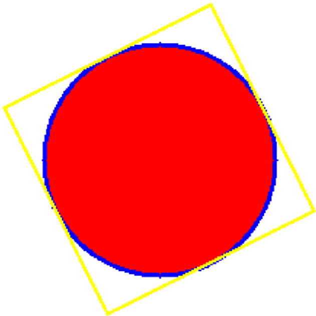

Figure 4 – *Visualization of the ellipse and the bounding rectangle demonstrated differences in symmetry and elongation between the varieties.*

The histogram of the color channels showed the dominance of the red component in Fuji and Brabourne varieties, while the green-yellow spectrum prevailed in Gold Delicious. Additionally, the proportion of pixels with a shade of red (Hue = 0-20°) was calculated, which made it possible to quantify the "redness" of fruits.:

Variety	Percentage of red (%)
Fuji	38,7
Brebourne	54,2
Hybrid	65,8
Gold Del.	12,6
Sinap Alm.	91,3

Thus, the color analysis in the HSV model showed a high diagnostic significance for varieties with pronounced coloration.

Comparison with traditional methods (visual inspection by an expert) has shown the following advantages of the proposed algorithm:

• increased processing speed by more than 5 times;
• elimination of subjectivity;
• possibility of automation of sorting in production conditions;
• Adaptability to new varieties with minimal customization.

Despite the high results, the system has limitations:

• Sensitivity to lighting (normalization of conditions is required);
• the possibility of overlapping features between similar varieties;
• Limitations on the analysis of deformed fruits.

In the future, it is planned to implement a real sorting stand with the connection of a weighing platform and integration with an accounting system (RFID, barcoding).

4. Conclusion

As part of the research, a method for automatically determining the apple variety based on the analysis of visual features obtained from digital images has been developed and experimentally tested. The proposed algorithm includes the stages of image capture, preprocessing, object selection, extraction of morphological and color features and classification based on statistical models (SVM, k-NN, logistic regression), without using neural network methods.

The experimental results showed that the proposed system is capable of providing an average accuracy of varietal identification at the level of 90%, demonstrating particularly high accuracy in classifying varieties with pronounced morphological or color characteristics (for example, Gold Delicious and Sinap Almaty). The analysis of shape (roundness, eccentricity) and color (hue, saturation, proportion of red) confirmed the diagnostic value of the interpreted features for the sorting task.

An additional advantage of the system is its reproducibility, low computational requirements and the ability to adapt to new varieties without the need for extensive retraining. This makes the proposed solution suitable both for laboratory analysis and for use in production sorting lines.

Future work is planned to focus on the integration of weight and texture characteristics, the expansion of the product range, as well as the creation of an industrial prototype portable sorting complex based on single-board computers with a user interface.

5. Gratitude

The article was prepared as part of research under a grant from the Ministry of Science and Higher Education of the Republic of Kazakhstan IRN AP19678983 "Development of digital technology and a small-sized machine for quality control and automatic sorting of apples into commercial varieties".

6. The list of references

1. 1. Nurtuleuov A., Moldazhan A., Kulmakhambitova A., Zinchenko D. Justification of the method and algorithm for determining the quality indicators of apples and their automatic sorting into categories // Scientific journal. - 2021. - No. 3 (91). - P. 125-133. DOI: 10.37884/3-2021/14.

2. Alikhanov J., Moldazhanov A., Kulmakhambetova A., Zinchenko D., Azizov A., Nurtuleuov A., Sarsenbekuly D. Digital technology for determining quality indicators and classification of apple fruits based on computer vision and deep learning // *AGRICULTURAL MACHINERY 2024. Proceedings of the XII International Scientific Congress*, 26–29 June 2024, Varna, Bulgaria. – P. 21–23.

3. 3. Jakhfer Alikhanov, Moldazhanov Aidar, Akmaral Kulmakhambetova, Dimitriy Zinchenko, Alisher Nurtuleuov, Zhandos Shynybay, Tsvetelina Georgieva, Plamen Daskalov Methodology for Determining the Main Physical Parameters of Apples by Digital Image Analysis Scientific journal "AgriEngineering" AgriEngineering Volume 7, Issue 3 March 2025 Article number 57 DOI: 10.3390/agriengineering7030057

4. Bhatt A.K., Pant D. Automatic apple grading model development based on back propagation neural network and machine vision, and its performance evaluation // *AI & Society*. – 2015. – Vol. 30(1). – P. 45–56. DOI: 10.1007/s00146-014-0540-7.

5. Sofu M.M., Erb O., Kayacan M.C., Cetisli B. Design of an automatic apple sorting system using machine vision // *Computers and Electronics in Agriculture*. – 2016. – Vol. 127. – P. 395–405. DOI: 10.1016/j.compag.2016.06.017.

6. Moallem P., Serajoddin A., Pourghassem H. Computer vision-based apple grading for Golden Delicious apples based on surface features // *Information Processing in Agriculture*. – 2017. – Vol. 4. – P. 33–40. DOI: 10.1016/j.inpa.2016.10.003.

7. Verdhan V. Computer Vision Using Deep Learning: Neural Network Architectures with Python and Keras. – Apress, 2021. – 420 p. DOI: 10.1007/978-1-4842-6616-8.