Detection of defects on apples using hyperspectral reflection visualization combining both vegetation index analysis and neural network

P V Balabanov, A G Divin, A S Egorov, A A Zhirkova and V A Yudaev
Tambov State Technical University, Tambov, Russia

E-mail: egorov.andrey@list.ru

Abstract. The article shows the possibility of using vegetation indices to build algorithms for recognizing defects in plant tissue of apples. The authors demonstrate that it is advisable to use halogen lamps as light sources for apples hyperspectral control, but LED backlighting to correct the spectrum at the edges of the 400-1000 nm range is preferable to use. The experimentally obtained spectra of the optical radiation reflected from the apples plant tissue in the mentioned wavelength range are presented and show the possibility of identifying defects using vegetation indices. At the same time, the risk of obtaining false defects for some varieties of apples is revealed. To reduce this risk, it is proposed to use artificial Deep Feedforward neural networks with one hidden layer.

1. Introduction
Apples are one of the most popular fruits among consumers all the world due to their taste and nutritional characteristics. In Russia, apples are the most accessible for the mass consumer fruits during the year, the annual production and consumption of which is growing steadily. Therefore, Company manufacturers and processing enterprises aim to automate the sorting process to improve the quality and to reduce the cost of apples processing.

Surface defects of single fruits affect the shelf life of the entire batch of products and its price for sale. The fruits of apple tree are susceptible to a wide range of diseases, and the automatic and rapid detection of defects is still a difficult task, despite the significant success of science in this field.

Detection of rotten, wilted, agricultural pest or phyto-diseases infected apples is traditionally carried out by visual inspection methods during the process of fruits sorting. Wherein, the first signs of rot on the surface of an apple are visually difficult to detect, this fact is explained by a relatively weak change in the color of the surface in the infected area. Therefore, in recent years, machine vision methods based on spectroscopy in the visible (Vis) and near infrared (NIR) regions have received significant development in the field of quality control of fruits and vegetables [1-3]. The advantage of spectroscopic methods for controlling fruit quality over other methods, for example, thermal imaging [4], radiographic [5], and magnetic resonance imaging [6], is high speed and relative cheapness.

The aim of our work is to create an information-measuring system for determining defects in apples transported on a conveyor, the principle of which is based on the of hyperspectral images processing. To achieve this goal, the following tasks were set: to determine spectrograms of the apples surface damaged by rot, phyto-diseases, agricultural pests; to analyze the spectrograms and identify the specific features inherent in the defects; to propose an approach to classifying apple defects according to their
spectrograms. At the basis of the developed system, we used hyperspectral control methods in Vis-NIR regions [1].

Spectral reflectivity is a characteristic feature of plants condition [2]. At the same time, the volumes of information about the control object obtained from hyperspectral cameras are very large, what requires large processing power and increases processing time [3]. Therefore, when creating high-speed sorting systems, it is necessary to reduce the volume of hyperspectral data [4]. The Principal component analysis (PCA) method [5] is one of the effective ways to solve this problem [6]. Another approach is the use of vegetation indices (VI), for the calculation of which only data obtained at specific wavelengths are used, but not a full range of hyperspectral data. Knowing the connection of VI with the qualitative condition of the fruit, it is possible to identify and to control defects [7]. The effectiveness of VI is determined by the features of reflection from the surface of objects, and the indices are determined empirically.

Several groups of the VI can be distinguished depending on their function. The first group includes indices (NDVI, SR, EVI, VOG1-3 and others), which reflect the content of chlorophyll in plant tissues [8]. Indexes are used to determine the presence and density of vegetation on the surface of the earth. In this group, the Normalized Difference Vegetation Index (NDVI) is most widely used. The calculation uses the Red and NIR spectral regions. This index is mainly used to assess the state of the crop in the fields, since it is sensitive to minor changes in the physiological state of plants. The indices of the second group (PRI, SIPI) make it possible to evaluate the efficiency of the light used by plants during photosynthesis. The growth rate of plants depends significantly on the efficiency of their carbon digestion, which allows using these indices to assess the illumination quality of plants grown, for example, in greenhouses. To indirectly determine the concentration of nitrogen in plant tissues, NDNI type indices are used and calculated using the mid-infrared range (SWIR). Since the nitrogen content in the soil significantly affects the growth rate of plants, the control of this index is important in the initial stages of vegetative development. To indirectly determine the carbon content in the form of lignin and cellulose, CAI, PSRI indices are used, taking into account the total amount of carbon in the form of lignin and cellulose, which present in wood and dry plant tissues in large quantities. Indexes CRI, ARI evaluate the content of the main pigments - carotenoids and anthocyanins, which appear when the fruits ripen or the vegetation is inhibited in significant quantities. To assess the moisture content in vegetation, indexes such as WBI, MSI, NDII are used.

There is a small number of works devoted to the determination of defects of apples based on index images. In [7], to identify sunburns and hailstones on the surface of apples, the CI and CIRE indices, respectively, we used and calculated from images at wavelengths of 678, 705, and 750 nm [8, 9]. At the same time, the presence of hailstones the CIRE index reflects better, which takes values of 0.3..0.4. A significant influence on its values is exerted by control object shading, which introduces errors in the results of defect identification. In [7], it was noted that damage accumulation is reflected by an increase in the content of anthocyanins, which allows the use of the corresponding mARI index. Thus, when determining defects on the surface of apples, it is potentially possible to use VI. For this purpose, it is necessary to obtain spectrograms of the apples surface, including uninjured and damaged areas, to calculate the indices that are most suitable for recognizing defects on the surface.

2. Materials and methods

2.1. Characteristics of the controllable objects

Apples of the 2019 harvest were used as objects of control. The varieties were following: “Oryol Sinap” - RF, Lipetsk Region, Lebedyansky District, “Agronom” State Farm (AS); “Orlik” - RF, Tambov Region, Petrovsky District, “Dubovoye” State Farm (AO); “Martovskoye” - RF, Tambov Region, p. Pervomaisky, “Snezhetok” company (AM). The volume of sample was at least 100 pcs of each kind of apples, the average fruit size was 50..70 mm. The color of the fruit is sporred yellow-green, red green. To obtain spectrograms of the apples surface areas, including the damaged area, we selected fruits
containing areas affected by scab, rot (area of at least 0.2 cm²), fruits damaged by agricultural pests (caterpillars of the apple moth).

2.2 Measuring the spectrogram of the control object
To obtain hyperspectral images, camera 1 (Specim FX10e) with a lens having a variable focal length is used (see figure 1). It is located at a height of 300 mm above the surface of the roller conveyor 4 on which the control object 3, illuminated by sources 2, is moving. The camera has a spectral sensitivity of 224 bands in the wavelength range from 400 to 1000 nm, and its sensor consists of 1024 sensitive elements with a resolution of 12 bits.

![Figure 1. The diagram of the technical vision system for apples hyperspectral quality control on a conveyor.](image)

For real-time processing of images from the hyperspectral camera, we used software installed on a PC computer and written in the Python programming language. The graphical user interface is created using the QT library. To determine the presence of the test sample on the conveyor, the reflection intensity \( R \) at a wavelength of 710 nm is analyzed (figure 1). The area with a larger value of the reflected radiation (in the example \( 210 < x < 920 \)) corresponds to the control object, the area with a lower value corresponds to the conveyor. To automatically determine the boundaries \( x_l, x_r \) of the control object, the coordinate \( x_{\text{max}} \) in the point of maximum intensity of the reflected radiation is determined and the average harmonic value \( R_g \) of the reflection intensities is calculated at all points located to the left and right of \( x_{\text{max}} \), after that the smallest of them \( (R_{\text{gmin}}) \) is selected. Graph points for which the condition \( R > R_{\text{gmin}} \) is satisfied correspond to the reflection intensity of the control object and \( R < R_{\text{gmin}} \) - for the conveyor.

Since only apples with defects exceeding the specified sizes are required to be rejected during the sorting process, it is necessary to calibrate the hyperspectral camera. As a result of calibration, the dependence

\[
l = \frac{p}{c}
\]

allows to convert the size of an object in image from pixels \( (p) \) to millimeters \( (l) \), using the known constant \( c = 6.49, \text{(px / mm)} \), which was found experimentally.

2.3 The choice of light sources for apples on the conveyor
The choice of light sources is an important task, since the sources must provide uniform radiation intensity in the spectral range of the used Specim FX10 camera. It is known that the most uniform spectrum has daylight, but it is difficult to use in vision systems. The spectrum of fluorescent lamps has pronounced peaks, which may lead to errors in determining the color of the object. The range of LED lamps eliminates the factor of excessive sensitivity to red light and compensates the weak sensitivity to blue and green. Considering this, in the range of 400..1000 nm, the intensities of reflection of various types of radiation from a flat surface were studied. The sources used were: halogen lamps (G) with a power of up to 100 watts, metal halide lamps with a power of up to 70 watts, xenon lamps (KS) with a
power of up to 35 watts, sodium lamps with a power of up to 70 watts, LED lamps (CB) with a power of up to 10 watts. All sources were equipped with directional reflectors, which made it possible to concentrate most of the radiation on the reflective surface, which was used as a sheet of matte white paper with a thickness of $104 \pm 2 \mu m$, CIE of $146 \pm 3\%$, brightness $R_{457}$ D65/10 $105\pm1,5\%$. The figure 2 shows graphs of the relative reflection intensity (NRI) dependence on the wavelength $W$, where NRI was determined by the formula:

$$NRI = \frac{R_I}{R_{I_a}}$$

where $R_{I_a}$ - the reflection intensity at the wavelength $W$ and at the average reflection intensity in the range of $400 \ldots 1000 \text{ nm}$.

![Figure 2](image)

**Figure 2.** The Intensity of the reflection from radiation sources.

3. **Investigation results and discussion**

One of the tasks facing us was to determine from the spectrogram the apple surface areas damaged by phyto-diseases. On the surface of the control object, we identified four areas (figure 3) - healthy tissue (GAP - good apple peel), containing a tissue defect (DAP - defect apple peel), containing a stalk (ASt - apple stem) and the opposite region containing the sepal (ASe - apple sepal).

![Figure 3](image)

**Figure 3.** The investigated apple surface areas.

To solve this problem, we used generally accepted in the world practice approach to assess the qualitative state of vegetation — the determination of the vegetation indices. To calculate the indices, spectrograms of apples, measured in the range $400 \ldots 1000 \text{ nm}$, were used. For apples of all three kinds AS, AO, AM (not depending on their color) the spectrogram of the GAP region contains a specific extremum at 680 nm (figure 4a), what indicates the presence of radiation absorption by chlorophyll in this region. For the scab DAP region, the extremum at 680 nm is negligible or absent (4b). That is, for healthy apple tissue, the spectrogram shows the presence of chlorophyll, and for damaged tissue, the absence. This feature may allow us to distinguish healthy or defective fruits by spectrogram.
The most common index reflecting the content of chlorophyll is NDVI. It is based on two sections of the spectral reflectance curve of vascular plants that are independent of other factors. The solar radiation absorption maximum of chlorophyll of higher vascular plants lies in the red region of the spectrum (0.6-0.7 μm). The maximum reflection region of the plant tissue with cellular structures lies in the infrared region (0.7-1.0 μm). The index was determined by the formula NDVI = (RI_750 - RI_680) / (RI_750 + RI_680) and had a range from 0.1 to 0.25 for the apple surface (figure 5) and from 0 to 0.05 for the conveyor area.

Moreover, for the area damaged by rot, the index value was less than for the undamaged areas. The similar pattern of the NDVI index changes was observed for areas containing the peduncle and sepals. Therefore, the use of NDVI makes it possible to distinguish the GAP regions from DAP, ASt, ASe, but it doesn’t allow us to distinguish the last three regions from each other. The last fact makes it difficult to make decisions about apples rejection based on specific changing the NDVI index in the control area.

When the fruit ripens, its color changes, due to the increase in pigments concentration - anthocyanins and carotenoids. Corresponding indices which allow us to estimate their content indirectly, were calculated using the formulas ARI\(_1\) = 1 / RI\(_{550}\) - 1 / RI\(_{700}\) and CRI\(_2\) = 1 / RI\(_{510}\) - 1 / RI\(_{700}\). The calculation results are presented in figure 6.

**Figure 4.** Spectrograms for healthy (a) and damaged (b) apple surface (AS).

**Figure 5.** NDVI index for AM: 1 - an apple damaged by rot; 2 - uninjured apple.
Both for the region of the peduncle ASt and for the defective region DAP, a decrease in the indices ARI1 and CRI2 is observed. This fact is explained by a significant decrease in the corresponding pigments (anthocyanins and carotenoids) in the apple plant tissue. The apple peduncle and the sepal region is characterized by the presence of woody and dry tissue. Therefore, in the indicated area, the index value $\text{PSRI} = (\text{RI}_{680} - \text{RI}_{500}) / \text{RI}_{750}$, often used in assessing areas with dry vegetation, increases slightly. For two varietes of studied apples (AS, AM), the type of these dependences is similar. The AO variety was an exception, because it is characterized by a predominance of red surface color and green stalk, what is the reason of the large extremum on the ARI1 plots in the ASt region, while the PSRI index, for most fruits, either constantly or slightly decreases (figure 7).

The results show that the nature of indices changing in various areas of the fruit depends not only on the type of investigated area (defect, peduncle, healthy tissue), but also on the apple variety. Probably it can be explained by the difference in surface color and moisture and solid components content in the plant tissue. This circumstance makes it difficult to obtain an universal empirical index that allows, by the nature of its change, to make reasoned decisions during the apples sorting process. However, for the ASt and DAP areas with various defects (rot, scab, worm), the nature of indices changing is different, what allows to classify the defects by using a neural network trained on spectrograms of preliminarily given apples variety.
4. Neural network for classifying apple defects

4.1. Input and output data, architecture
To solve the classification task, we chose the Deep Feedforward neural network with one hidden layer, which architecture is shown in the figure 8.

![Neural network architecture](image)

**Figure 8.** Neural network architecture.

To train the neural network, we used 10,000 spectrograms for the areas of GAP, DAP, AST, ASe. The size of each area was 2 mm or 12 points. At each point, the values of the indices NDVI, PSRI, AR11, CRI2 were calculated. Thus, the number of neurons of the input layer (Input cells) is $12 \times 4$. The number of outputs is determined by the number of defect classes. The following classes and corresponding numerical codes were identified by the presence and nature of defects: GAP (0), a wilted apple (1), an apple with rot (2), an apple with mechanical tissue damage (3), a worm apple (4), an apple with scab (5), ASe (6), AST (7). Thus, the output layer (Output cells) consists of 8 neurons (corresponds to one of eight classes of the control object). The number of neurons in the hidden layer (Hidden cells) was selected experimentally and is equal to 64. The sigmoid was selected as the activation function in the hidden layer.

4.2. Testing results
Neural network testing was carried out on 5000 spectrograms. Moreover, the number of errors depends on the number of training epochs (figure 9).

![Dependence of the error quantity on the number of epochs](image)

**Figure 9.** Dependence of the error quantity on the number of epochs.
It is necessary to note that the previously shown results provide the possibility of using the neural network trained on the apples index images in order to recognize their defects.

5. Conclusion
Thus, approaches for the development of an information-measuring system (IMS), which allows to determine the surface defects for apples transported on a conveyor, are considered in the presented article.

The authors showed the possibility of using the vegetation indices (VI) to identify apples damages resulting from phyto-diseases. This approach is quite possible for some varieties of apples, but it is not universal, since errors of the first kind are possible due to the acceptance of the AS稳定 and ASe regions of the fruit plant tissue as a defect. The way out can be the use of an artificial neural network trained on a large number of apples containing predetermined defects, including the AS稳定 and ASe areas. This conclusion is based on the fact that the spectra for defective plant tissues visually differ from the spectra of healthy sites.

The presented IMS can become in future the main element in robotic sorting of apples, as well as other fruits and vegetables.

Acknowledgments
The work was performed using the equipment of the “Robotics” collective-use center (registration number 677349) of the Tambov State Technical University. This work was supported by the Ministry of Science and Higher Education of the Russian Federation in the framework of the project under Agreement No. 05.604.21.0240, unique identifier RFMEFI60419X0240.

References
[1] Lu Y and Lu R 2018 Structured-illumination reflectance imaging coupled with phase analysis techniques for surface profiling of apples J. Food Eng. 232 11–20
[2] Li J, Chen L, Huang W, Wang Q, Zhang B, Tian X, Fan S and Li B 2016 Multispectral detection of skin defects of bi-colored peaches based on vis-NIR hyperspectral imaging Postharvest Biol. Technol. 112 121–33
[3] Huang W Q, Li J B, Wang Q Y and Chen L P 2015 Development of a multispectral imaging system for online detection of bruises on apples J. Food Eng. 145 62–71
[4] Baranowski P, Mazurek W, Wozniak J and Majewska U 2012 Detection of early bruises in apples using hyperspectral data and thermal imaging J. Food Eng. 110(3) 345–55
[5] Haff R P and Toyofuku N 2008 X-ray detection of defects and contaminants in the food industry Sens. Instrum. Food Qual. Saf 2(4) 262–73
[6] Zhang L and McCarthy M J 2012 Black heart characterization and detection in pomegranate using NMR relaxometry and MR imaging Postharvest Biol. Technol 67(5) 96–101
[7] Solovchenko A E, Lukyanov A A, Nikolenko A A et al. 2020 Physiological foundations of spectral imaging-based monitoring of apple fruit ripening Acta horticulturae
[8] Merzlikin M N, Solovchenko A E and Gitelson A A et al. 2003 Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit Postharvest Biology and Technology 27 197-211
[9] Gitelson A and Solovchenko A 2017 Generic algorithms for estimating foliar pigment content Geophysical Research Letters 44 9293-8