

An approach for eliminating restrictions on the synthesis of quality classifiers of fruits and vegetables

Radoslava Nikolova Gabrova^{*}, Lena Filipova Kostadinova-Georgieva

Dept. Computer systems and technologies, University of Food Technologies, Plovdiv, Bulgaria

Email address:

rgabrova@yahoo.com (R. Gabrova), lenakostadinova@yahoo.com (L. Kostadinova-Georgieva)

To cite this article:

Radoslava Nikolova Gabrova, Lena Filipova Kostadinova-Georgieva. An Approach for Eliminating Restrictions on the Synthesis of Quality Classifiers of Fruits and Vegetables. *International Journal of Science, Technology and Society*. Vol. 2, No. 5, 2014, pp. 109-114. doi: 10.11648/j.ijsts.20140205.13

Abstract: Synthesis of classifiers of agricultural products by quality is related with two main tasks –selection symptoms for object description in new space and choice of method for pattern recognition for objects separation in quality classes. A special feature of agricultural products as objects of classification is the different size of their primary descriptions. This feature imposes restrictions on the choice of symptoms and a method for pattern recognition in the synthesis of the classifier. The work proposes an algorithm for unification of the primary descriptions of classified products using interpolation methods. Different methods of interpolation are compared and the one which provides the simplest algorithm is recommended for use. The algorithm is applied to a virtual extension of experimentally derived primary descriptions of potato tubers. The new extended descriptions are applied to the synthesis of the symptoms and classifier. Simulation testing of the classifier, synthesized with the new and original descriptions of the products was conducted in Matlab. The applicability of the proposed algorithm to unify the descriptions of classified agricultural products is proved. The proposed approach removes the restrictions on the choice of method for the synthesis of symptoms and method for pattern recognition and reduces the number of training set of objects.

Keywords: Qualification of Agricultural Products, Interpolation, Classifier, Pattern Recognition

1. Introduction

In optical systems for classification according to quality (automatic qualification and sorting) of agricultural products, receiving physical characteristics of objects, namely for the current state of their quality, includes two functions: passing objects to the measuring position and opto–electronic conversion of the optical "images" of the product. Based on the reflected from or transmitted through the product light, an image of the product, one–dimensional or two–dimensional, representing the distribution of brightness is formed. In opto–electronic converter luminance signal is converted to one–dimensional electrical signal function of time.

The process of the targeted treatment of the resulting primary information for classifying products into quality classes can be divided into the following stages [2, 3, 5]:

- Pretreatment of primary information about the object – includes classic operations of converting signals as normalization, sampling and quantization (analog to

digital conversion), filtration, preliminary computational procedures etc.

- Transforming the images of objects in another (new) symptom space – the description of the images are formalized by forming symptoms for recognition;
- Identification of the product belonging to a particular quality class – a chosen method of pattern recognition, correlates symptoms of qualified images and reference product descriptions of the classes with different qualities and based on the corresponding algorithm a decision is made for belonging of an object to one of the defined classes.

In accordance with the described features, automatic system for qualifying and sorting of agricultural products can be considered an information system, requiring coordination of information flows between functions [4, 6].

The agricultural products have different geometric dimensions and electrical signals (functions of time), formed from the primary received information about the current state of the quality of individual products, will be of varying duration. As a result in the subsequent operation of

digitization (sampling and quantization) of these electrical signals a description of objects with different number of reported values is obtained. This feature can be a disturbing factor in the coordination of information flows in subsequent operations as synthesis of symptoms and implementing variety of methods for pattern recognition in the classifier.

To eliminate this inconvenience the method of synthesis of symptoms based on heuristic synthesized relationships can be used [1]. As symptoms for recognition are used, maximum value, minimum value, average value, a maximum adjacent increments etc.. Synthesis of such symptoms significantly depends on the subjective human factor – the knowledge, experience and intuition of researchers. A disadvantage is that the synthesized symptoms are often correlated with each other and their information value cannot be assessed adequately.

Another approach for synthesis of symptoms is based on rigorous mathematical (theoretical) methods for the formation of the symptoms space [7]. One direction of this approach is the synthesis of symptoms, based on the spectral methods of functional analysis. Various orthogonal decompositions – Fourier, Cosine, Hadamard and Hartley are used. The coefficients of the selected transformation are used as symptoms for recognition. The spectral components are uncorrelated and those with smaller numbers contain a large proportion of the signal energy. Therefore, the symptoms for recognition (coefficients of decomposition) are uncorrelated and those with smaller indices will be more informative.

The coefficients of the discrete Fourier transform (DFT) $S(n)$ ($n = 0, 1, \dots, N-1$) are determined according to [10]:

$$S(n) = \sum_{i=0}^{N-1} s(i)e^{-j2\pi ni/N}, \quad n = 0, 1, \dots, N-1, \quad (1)$$

where $s(i)$ ($i = 0, 1, \dots, N-1$) is the initial description of an object with the N number of points. They give a description of the object in the new symptoms space. In accordance with (1), one of the most informative symptoms – the zero coefficient of the decomposition $S(0)$ is determined by:

$$S(0) = \sum_{i=0}^{N-1} s(i). \quad (2)$$

When descriptions of the studied objects have different sizes the sum of (2) will contain a different number of addends and therefore the values of $S(0)$ for different products of the same class will vary substantially. Therefore, the deviation in the size of the descriptions of the studied objects is a disturbing factor in the selection of the spectral group of methods for synthesis of symptoms and requires adaptation of these methods. This in turn may lead to a requirement for adapting algorithms of different methods for pattern recognition.

Such reasoning on the influence of differences in the size of the descriptions of the products can be attributed to other

methods for synthesis of symptoms and methods of pattern recognition.

In this sense, the task of object recognition with different size of the descriptions imposes restrictions on the choice of methods for synthesis of symptoms and/or pattern recognition or requires their adaptation.

The work proposes an approach in which is applied prior image processing, and thus is eliminated the mentioned disturbing factor, allowing the application of all methods for synthesis of symptoms and methods for pattern recognition without restrictions. This preliminary image processing consists in the uniformity of their size using the method of interpolation from the theory of "Numerical Methods".

2. Materials and Methods

The initial description of an agricultural object under study is represented by vector (sequence of numbers) with a fixed length and the measured values of signal in the reported points are its coordinates:

$$s = \{s(0), s(2), s(3), \dots, s(i), \dots, s(N-1)\}. \quad (3)$$

Each of the reported values $s(i)$ contains information about the quality of the product in a certain fraction of its total volume.

To eliminate confounding factors (deviation in the size of descriptions) in the synthesis of quality classifiers of agricultural products the proposed algorithm is setting a new description of each object with a number of reporting points equal to the maximum number of diskrets from the descriptions of all the objects of general aggregate. This new description is derived based on the experimental data for the object by increasing the number of diskrets by defining intermediate points using the method "interpolation".

Adaptation of the 'interpolation' from the theory of "Numerical Methods" to the task of classification of agricultural products makes it possible to align the size of the object descriptions. Since interpolation is a method of constructing (calculation) new numeric values between known ones, it can be concluded that the new obtained values of the image of the object will contain reliable information on the quality, as it is defined on the basis of existing reporting points, containing such information.

Interpolation algorithm comes down to finding a mathematical description of the relationship (function) passing through the existing points (interpolation nodes) and by using this dependence to determine (calculate) new "intermediate" points.

For the purpose of the interpolation can be used interpolation functions from different classes, for example [8, 9]:

- algebraic polynomial

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0; \quad (4)$$

- trigonometric polynomial

$$T(x) = \frac{a_0}{2} + \sum_{k=1}^n a_k \cos kx + b_k \sin kx ; \quad (5)$$

- exponential polynomial

$$E(x) = a_1 e^{\alpha_1 x} + a_2 e^{\alpha_2 x} + \dots + a_n e^{\alpha_n x} \quad (6)$$

and others, where x is the independent variable, a_i and b_i are coefficients of the polynomial.

Of mentioned functions for interpolation most commonly are used those with an algebraic polynomial approximation, since these polynomials are calculated easily and the theory is best developed. Classical interpolation with given n points conditionally enrolled (accepted) coordinates x_i and y_i ($i = 0, 1, \dots, n-1$) in the plane is the definition of the interpolation polynomial $P(x)$, of $(n-1)^{\text{st}}$ degree, the values of the function at the interpolation nodes are equal to the default values – $P(x_i) = y_i$.

To determine the coefficients of the polynomial on the basis of (4) may be composed linear system of n equations (with n unknowns) of the form:

$$P(x_i) = y_i = a_{n-1} x_i^{n-1} + a_{n-2} x_i^{n-2} + \dots + a_1 x_i + a_0.$$

Another form of representation of the approximate polynomial function can be written using Lagrange polynomial:

$$L_n(x) = \sum_{j=0}^{n-1} y_j l_j(x) \quad (7)$$

where

$$l_j(x) = \prod_{\substack{k=0 \\ k \neq j}}^{n-1} \frac{x - x_k}{x_j - x_k} . \quad (8)$$

It should be noted that after the conversions the same record as (4) is obtained and the difference lies in how to determine the coefficients.

Forms of representation by Newton polynomial or polynomial Ermita can also be used. For drawing in both cases, however it is necessary to assign values of the derivatives in the nodes, i.e. to set the finite differences.

Since the deviations of the approximate function of the original curve can be significant in a large number of interpolation nodes and therefore higher order polynomial (the order of the polynomial is defined as from the number of interpolation nodes is removed unit), in practice often is used so-called piecewise polynomial interpolation. In this approach, the interval is divided into subintervals, where the data is interpolated with polynomials of lower degree that is well coordinated in the common edges of the subintervals.

Some of the varieties of piecewise polynomial-interpolation are:

- Piecewise constant interpolation or nearest-neighbor interpolation – the value of the unknown intermediate point equates the value of the closest standing interpolation node. The values of the interpolated points in the range $((x_{i-1}+x_i)/2, (x_i+x_{i+1})/2)$ are replaced with the values of y_i (the values of the function for x_i), i.e. in this range the interpolation function is a constant (0th order polynomial). A break obtained throughout the interpolation function in the interpolation nodes;
- Piecewise linear interpolation – it is assumed that the interpolation relationship in subintervals can be described by a linear function (polynomial of the first order);
- Piecewise cubic interpolation – it is assumed that the relationship in subintervals can be described by a cubic function (a polynomial of third order). This approach ensures equality of zero and first derivative of the given and an interpolated function, hence resulting in a smoother curve.

When the interpolation nodes are a large number and the target is a smooth functional relationship the method of interpolation by spline functions may be used. This approach assumes that the dependence of subintervals is a polynomial function of lowest possible order. This approach ensures the continuity of the derivatives of the approximate function to equal the order of the spline used minus one. Most often is used the spline of third order, i.e. relationship in subintervals is described by a cubic function and continuity of zero, first and second derivatives is ensured.

Fig. 1 shows graphically an exemplary description of 12 reports of an investigational object (blue) and interpolated description with 25 points (red), respectively using the following interpolation methods: a) Piecewise constant interpolation, b) Piecewise linear interpolation, c) Piecewise cubic interpolation and d) Spline interpolation.

Each point of the newly interpolated images has adequate information about the quality of the product, as it is determined, based on the information available for this quality condition.

Using Piecewise constant interpolation new values are not received but only some of them are repeated several times. This interpolation can not amend the maximum, minimum, etc. In fact, this interpolation is implemented most easily.

Increasing the order of polynomial interpolation often only complicates the algorithm. As it is seen from the figures, interpolation by Piecewise linear interpolation, Piecewise cubic interpolation and Spline interpolation gives similar results – keeps the trend of variation of the curves with very slight deviations of actual values from interpolated even some of them overlap.

On X-axis is recorded the number of the diskreets and physical can not be fractions, but fractions are obtained for most of the interpolated points. Therefore, the resulting interpolated description is translated so that the arguments

of the interpolated points are replaced by whole units from one to the maximum number of reports in the descriptions of specific experimental data. In Fig. 2 are shown graphic images of the object (blue), interpolated (red) and interpolated translation description (black) of the model object from Fig. 1 for Piecewise linear interpolation.

Physically realized process of interpolation, in order to align the size of descriptions of various objects meet the extension of virtual images of products and can be

interpreted as the use of different sampling step at their primary obtained description.

After applying the algorithm the descriptions of all objects contain an equal number of reports at the maximum possible number of specific experimental data, i.e. descriptions of all objects are the same length.

Descriptions of the objects can be extended to set a larger number of reports of the maximum of the experimental data.

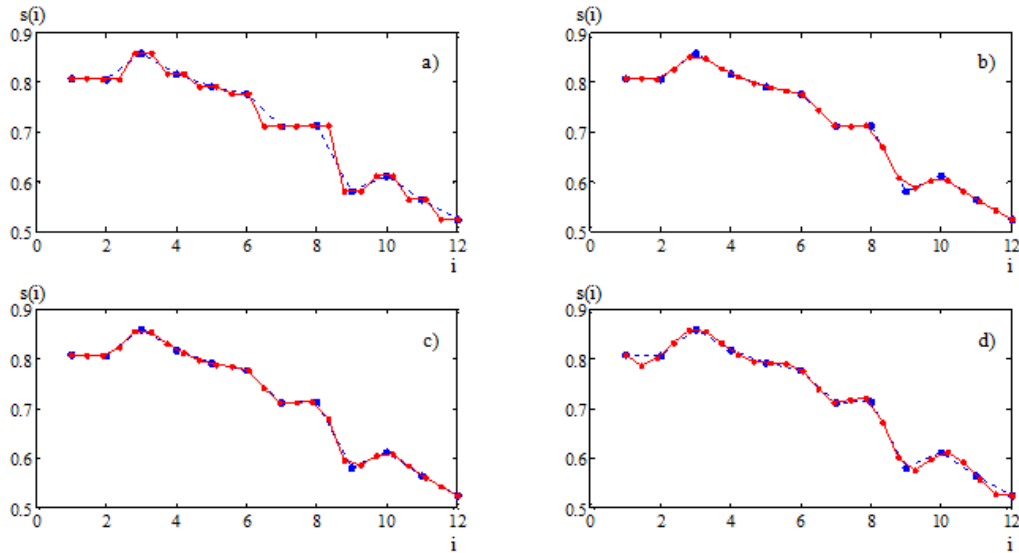


Figure 1. Diagram of an exemplary description of 20 reports of investigational object (blue) and interpolated description with 25 points (red), respectively, using methods of interpolation: a) Piecewise constant interpolation; b) Piecewise linear interpolation; c) Piecewise cubic interpolation; d) Spline interpolation.

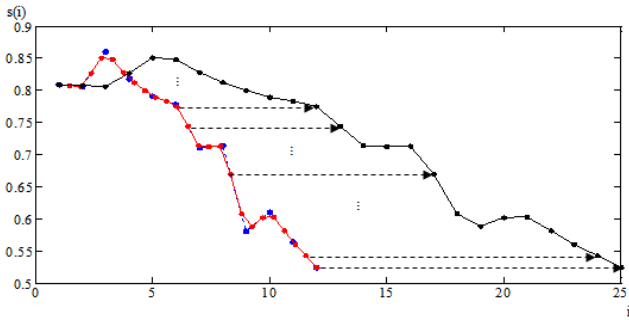


Figure 2. Images of the object (blue), interpolated (red) and interpolated translation description (black) of the model object for Piecewise linear interpolation.

During the synthesis of classifier the algorithm for aligning the descriptions by interpolation has to be applied to the entire set of experimental data, based on which is formed the so-called training and control set of objects, and in real application of classification – to the incoming objects for classification.

3. Results and Discussion

For verification the performance of the adapted method of pattern recognition are used initial descriptions of whole, unpeeled potato tubers of variety "Agria" [1]. The data for the values of light transmittance are experimentally obtained as the ratio of the signals from the photo-electronic sensors at two different wavelengths λ_1 and λ_2 and longitudinal scan of products, classified by experts in three classes. Descriptions of individual objects contain different numbers of reported values of the light transmittance recorded at scanning. Description with the shortest length contains 9 readings and with the greatest length – 25. Table 1 shows the number of objects in each class and the corresponding number of reported values that the description contains, i.e. objects are grouped by the length of their description.

Table 1. Number of objects in each class and the corresponding number of reported values that the description contains.

Classes	Number of reported values																
	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 st class	0	1	2	12	34	75	107	153	208	159	92	87	80	59	29	21	26
2 nd class	3	4	5	8	15	21	88	148	188	168	133	112	73	46	57	45	55
3 rd class	0	0	1	8	3	13	48	83	133	104	103	95	64	78	67	45	35

Of the total general set of objects, at random and without a second choice are formed three different training set of objects, containing respectively 60, 120 and 180 objects from each class. The rest multiple objects of each class is used as a control sample.

Classifiers have been tested with the original descriptions of the objects and with interpolated descriptions by the methods Piecewise constant interpolation, Piecewise linear interpolation, Piecewise cubic interpolation and Spline interpolation. In the process of testing the heuristic synthesized symptoms, method of pattern recognition "distance to the reference" and the various training set of objects are used. Interpolated descriptions contain the maximum number of points for specific reporting experimental data, which are 25.

The parameter "error of classification for a class" that evaluates incorrectly classified objects in the class is used for evaluating the performance of the classifiers. The two class classifier has two types' errors – error of higher quality class and error of lower quality class. Errors due to transition from the class with lower quality to that with higher, according to the theory of statistical decisions, can be classified as errors of second kind (omission of the

target) and errors due to the transition from best to worst class – as the error of first kind (false alarm). In determining the errors an approach to compare the classes against each other was used. In the case of three classes' classifier, the important classification mappings are defined—first class against second class, first class against third class, second class against the third class. For presenting the results of the comparison of class i against class j (i / j) are used designations:

- $F_{i/j}^I$ – the error of first kind;
- $F_{i/j}^{II}$ – the error of the second kind.

The results from the simulation in Matlab shown in Fig. 3 are identical for the actual and interpolated descriptions by all used interpolation methods and the various training set of objects. Therefore the choice of a particular method of interpolation should depend only on its complexity. Actually Piecewise constant interpolation method is realized most easily and gives the same results as other interpolation methods, applied as pre-treatment of descriptions of the objects in the synthesis of classifiers therefore it may be preferred over others.

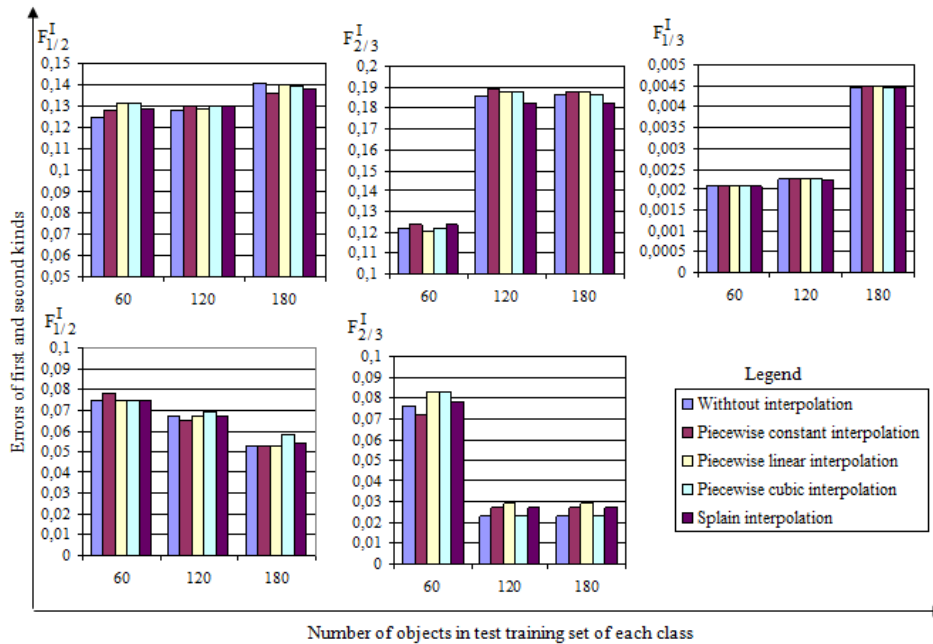


Figure 3. Errors of the first and second kind for the studied classifiers in the original descriptions of objects and interpolated descriptions for methods Piecewise constant interpolation, Piecewise linear interpolation, Piecewise cubic interpolation, Spline interpolation, heuristic synthesized symptoms and method of pattern recognition "Distance to the reference" and the various training set of objects.

4. Conclusion

In obtaining the initial information of the quality of agricultural products by scanning the entire length and subsequent digitalization, descriptions of objects of different lengths are recorded. This feature appears to be a disturbing factor in the synthesis of classifiers of agricultural products, concerning the choice of method for

synthesis of symptoms and method of pattern recognition.

An approach for unification the description of the agricultural products, by adapting the method of interpolation is proposed. The application of the suggested interpolation algorithm is an equal number of reports of the descriptions of all objects at the maximum possible number of specific experimental data.

Four different interpolation methods are compared. From the viewpoint of ease of realization of the algorithm

Piecewise constant interpolation may be preferred.

The algorithm is applied to a virtual extension of experimentally derived primary descriptions of potato tubers. The new extended descriptions are applied to the synthesis of the symptoms and classifier. Simulation testing of the classifier, synthesized with the new and original descriptions of the products is conducted in Matlab. The applicability of the proposed algorithm to unify the descriptions of agricultural products is proved. The proposed approach removes the restrictions on the choice of method for the synthesis of symptoms and pattern recognition and reduces the number of training set of objects.

References

- [1] A. Georgiev, L. Kostadinova, Automated systems for qualification and sorting agricultural products, Journal "Food Industry" №6, Sofia, Bul, 1985.
- [2] A. Georgiev, L. Kostadinova, R. Gabrova, On-line Sorting Apples in Modern Technologies for Providing Quality of Fruit and Vegetable in Chains, Acta Horticulturae, N 712, Proc. Of the IV Intern. Conf. on Managing Quality in Chains, vol. 2, p. 911–914, MQUIC2006, Bangkok, Thailand, 2006.
- [3] A. Georgiev, L. Kostadinova, R. Gabrova, N. Shopov, Increasing the Rate of Operation of Automatic Quality Classifiers for Agricultural products – Software and Hardware Decisions, Acta Horticulturae, N858,p. 431-438, March, 2010
- [4] I. Witten, F. Eibe, Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2005.
- [5] L. Kostadinova, A. Georgiev, R. Gabrova, N. Shopov, Improving the Automatic Classifiers in Systems for on-line "Tracing" the Quality of Agricultural Products, Proceedings Vol. 1 of UNITECH'07. Gabrovo, p.465-469, Bul, 2007.
- [6] L. Kostadinova, A. Georgiev, N. Shopov, Informational Aspects of the Processes in Objectively Determining the Quality of Food, , Proceedings Vol. 1 of UNITECH'10. Gabrovo, p.I-548-I-553, Bul, 2010.
- [7] N. Shopov, A. Georgiev, L. Kostadinova, R. Gabrova, Generating symptoms' spaces by orthogonal basis functions during automatic qualification of products, Proceedings Vol.1 of UNITECH'07, Gabrovo, p. 420-425, Bul, 2007.
- [8] R. Burden, D. Faires D, Numerical Analysis 9th Brooks/Cole, USA, 2011.
- [9] S. Chapra, R. Canale, Numerical methods for Engineers 6th ed., McGraw–Hill Higher Companies, USA, 2010.
- [10] S. Donevska, B. Donevski, Fast Fourier Transform, Sofia, Bul, 1999.