**SZENT ISTVÁN UNIVERSITY**

INNOVATIVE CLASSIFICATION OF THE CULINARY
PAPRIKA WITH AN AUTOMATIC IMAGING SYSTEM

Thesis of the doctoral (PhD) dissertation

Zoltán Gergely

Gödöllő – Hungary

2016

## CONTENTS

# LIST OF SYMBOLS

| | | |
|---|---|---|
| $\gamma$ | relative angles of curvature | [°] |
| $\varphi$ | absolute inclination angles of the lines | [°] |
| $EH$ | front to back factor | |
| $g$ | degree of curvature (factor) | [-] |
| $h$ | length of the paprika (without stalk) | [mm] |
| $h_{kköz}$ | approximate centreline length of the paprika | [mm] |
| $h_t$ | the total length of the paprika (with stalk) | [mm] |
| $m$ | distance of the end points of the segment appointing the line of the stem, measured on the $x$ axis | |
| $M_{1,2}$ | points marking the paprika's shoulder | |
| $P_1$ | point marking the tip of the paprika | |
| $P_{2,3}$ | points marking the paprika's stem | |
| $P_4$ | first major contour point of the paprika | |
| $\overline{P_4}$ | last major contour point of the paprika | |
| $S$ | the gravity centre of the paprika formation | |
| $v$ | shoulder width (shoulder diameter) | [mm] |
| $\hat{v}$ | iteration factor of the modified shoulder width (shoulder diameter) | [mm$^2$] |

## 1.    INTRODUCTION, OBJECTIVES

### 1.1.    Relevance and Significance of the Topic

Due to market needs and new legal regulations there is a strong demand for the qualitative classification of various types of produce in agriculture. The automation of such classification, however, requires new procedures and methods. These, in turn, require scientific elaboration.

Hungarian culinary paprika is a so-called Hungaricum (a Hungarian speciality of high value and representative of our culture), which is very popular both in Europe and Asia. This is also promoted by the fact that the competent organs of the EU have classified the paprika grown in Szentes as a type with a geographical indication under protection, within its food quality marking system.

In most parts of the world, the bell pepper (a so called blocky type) is known as pepper or paprika, which has long had its automatic classification systems. Most of these classification procedures, however, are based on weight measurement and therefore are capable of classifying round types of produce (tomatoes, potatoes, bell pepper, apples, apricots, cantaloupes, lemons, oranges etc.), and cannot be used for conic or pointed produce. This means there is no such automated classification procedure for the tapering white paprika that would meet all of the existing requirements. Since the demand for packaged culinary paprika has increased (primarily due to supermarket chains), the need for automated classification is rising too.

Even though technological advancement (especially in the field of electronics, automation and IT) makes paprika's classification possible, the lack of high-speed shape recognition, noise filtering and elaborate decision-making mechanisms, conic types of paprika are still sorted manually, both in high- and low-volume production plants. As a result, paprikas are brought to the market in sizes assigned by their respective producers that are not always in line with regulatory or market requirements. Classifying paprika also requires personnel with proper qualifications and routine, which leads to difficulties in organizing work, therefore it is of primary importance to lay the scientific foundations for an automated classification procedure.

### 1.2.    Objectives

Based on the market situation, the regulatory background and our technological advancement, there is a real and satisfiable need for an automated classification of conic and pointed types of culinary paprika.

The major goal of my research paper is to provide such a solution for these types of paprika that makes the classification procedure fast, accurate and in

line with all the regulatory requirements. Once the theoretical foundations of such automated classification are laid down, a next objective is to build an actual sorting equipment for running all the processes to identify the proper model and procedure of classification.

In building the required highly reliable machine, I plan to use an embedded measuring and processing system that ensures real-time, automatic classification at low cost.

To achieve that, the following tasks are to be completed during the project:

- Reviewing procedures capable of image segmentation; adapting these to embedded systems; and preparing proprietary algorithms if necessary.
- Building a measuring system with embedded technology in order to be able to start off with the measurements; integration of algorithms.
- Elaborating the method for the physical description of conic and pointed types of paprika, as well as developing the classification algorithm.
- Examining the camera and lighting system and checking the accuracy of the segmentation algorithm.
- Creating a real system of automated classification; examining its operation; analysing the present manual system and the created automatic system; as well as identifying the processes based on the finished automatic system.
- Running experiments concerning the general applicability of the elaborated procedure.

## 2. MATERIAL AND METHOD

In this chapter, I will present the parameters and framework conditions that determine the scope of my scientific project. I will also describe the types of paprika to examine, as well as the methods and technologies to be applied during classification.

### 2.1. Types of Paprika Examined

In accordance with market demands, the following types of mass-produced conic and pointed paprika have been examined:

- Conic types of paprika for stuffing (mostly of white colour),
- Pointed types (mostly sharp in taste),
- Special types with conic or pointed shape: kapia, "dolce italiano" and "corno di toro".

Since the market share of special types (so-called choice-widening types) is continuously growing, my research pays special attention to the analysis of these types. However, types with bell or tomato shape are not the subject of this research, since the automatic classification of these types of peppers is already solved due to their round shape, with machines using weight measurements.

The scale of examined paprika types is rather wide, comprising varied harvesting conditions. White and pointed types are usually harvested at their so-called economic ripeness (fully developed, hard fruit, shiny surface), whereas kapias are picked at different states, but mostly when they are biologically ripe (fully ripened state, with a red or yellow coloured fruit, depending on the species). This means that the colour of the paprika has been a crucial parameter from



Figure 2.1 Types of paprika ripening from white or greenish white to red, for stuffing (from left to right); HRF $F_1$, Hó $F_1$, Cecil $F_1$, Hajdú $F_1$, Creta $F_1$, Julianus $F_1$, Galga $F_1$, Flexum $F_1$, Emese $F_1$, Jász $F_1$
(Photos: Variety Maintainers)

the point of view of the optic system to be developed. In the samples there were both economically and biologically ripe pieces of paprika. My experiments involved conic and pointed paprikas with varied characteristics (shape and fruit colour) from the types shown in figures 2.1 and 2.2.

As is clear from the pictures, conic and pointed types of paprika are very diverse as to their colour and shape, and this demanded special attention during the development and planning of the optic system as well as of the image processing and classification algorithm.



Figure 2.2 Specialty paprika types ripening from green to red/yellow (from left to right); Karamida $F_1$, Kárpia $F_1$, Mágus $F_1$, Canal $F_1$, Red Marconi, Golden Marconi, Uranus $F_1$, Titán $F_1$, Rapires $F_1$, Kard $F_1$
(Photos: Variety Maintainers)

## 2.2. Developing the Measuring System

The shape recognition and analysis system had to be fit in the actual classification process without slowing it down or influencing it adversely, and without causing mechanical damage to the fruit. As a first step, the present process of classification had to be reviewed, which is shown in Figure 2.3. Together with the shape recognition system, all other parts of the classification system were elaborated, but these are not the subject of the present doctoral thesis.

The major parts of the measuring system are the following:

- Feeding and directing system
- *Shape recognition optic system:*
  - *CCD,*
  - *CCD processing unit,*
  - *optic system,*
  - *light sources,*
  - *data processing and assessment (classification) module.*
- Driving regulation system
- Fruit selection system
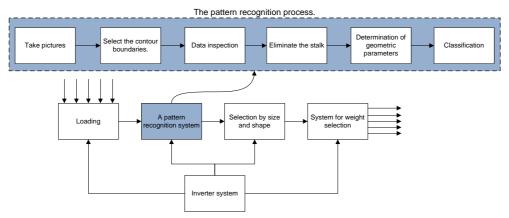- Equalizing system (based on weight)

Figure 2.3 The place of shape recognition within the process of classification

One characteristic feature of the classification process is that the machines must work on a periodical basis, as the produce ripen, but in these periods they must operate with high reliability, since any outage may result in significant damages. In order to reach a proper and competitive classification performance, the equipment must be able to run the full recognition process for 3-4 pieces of paprika per second. This requires a rather high-speed processing system.

Using more than two optic systems did not result in significant improvement in accuracy, but the complexity of control and synchronization, as well as the problems presented by the increased amount of data made it clear that this option should be rejected.

## 2.3.    The Developed Equipment

The separate parts of the system necessary for the classification had to be fit in the complete equipment carrying out the whole process.

For the reliable operation of image processing, for example, the role of the feeding module is very important, since this is the part that ensures the proper and even flow of produce on the line under the cameras. The individual pieces of paprika are arranged in the proper direction, but they must be over sighted by human personnel in order to remove any pieces that might stain or otherwise cause harm to the equipment. Since the elevator has rollers, it does not only move the peppers upward, but also turns them around, thus making the job of pre-selection for the personnel easier - the individual pieces do not have to be taken in the hand and turned around manually in order to be examined. In case there happen to be too many faulty pieces at the same time, the personnel can slow down the speed of the sloping elevator with a pedal, so they can continue to check all the items properly and remove any faulty pieces.

The individual peppers are sorted into classes by a system using condensed air. When the peppers get to their respective gates, an electrically controlled valve will make the nozzle blow them off the sorting line with approximately 6 Bars to the collecting boxes. These boxes are the Hungarian standard M30 boxes used in most horticultures. The even filling of the boxes is ensured by spring-strained carts moving on a sloped pathway, and these carts are equipped with weight measuring units. Figure 2.4. contains photo of the installed experimental system.



Figure 2.4 The prototype system in operation (Photo: Anna Tömpe)

### 2.4. The Image Sensing Process

The block diagram of the complete image sensing process can be seen in Figure 2.5. In simple terms, the operation of the system is as follows: the router sends a signal to the cameras at a rate required by the speed of the conveyor belt so that the perpendicular frames can be prepared at the right moments (sampling). After this, the data are sent to the central processing unit, whose job is to assess them and sort the peppers into classes based on the preliminary set of rules, as well as to create the intervention signal for the module responsible for the sorting.
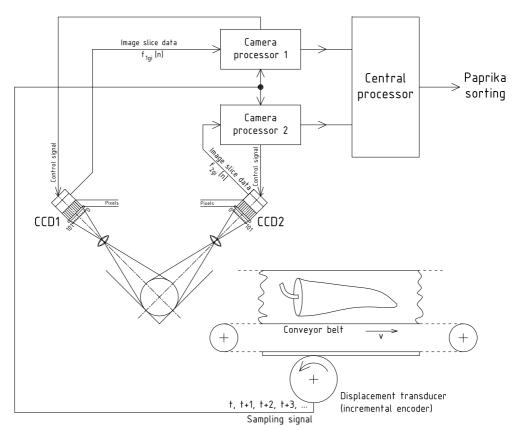
Image slice data
$f_{1gi}[n]$

Camera processor 1

Central processor

Paprika sorting

Camera processor 2

Control signal

Image slice data $f_{2gi}[n]$

Control signal

Pixels

Pixels

CCD1

CCD2

Conveyor belt

v

t, t+1, t+2, t+3, ...

Displacement transducer (incremental encoder)

Sampling signal

Figure 2.5 Schematic design of the image sensing process (without the illumination)

## 2.5. Determining the Geometry of the Paprika

Based on requirements, the most important parameter for classification is the shoulder width (shoulder diameter) of the paprika, which is practically speaking the maximum diameter of the fruit. This parameter is complemented by measuring the length, which ensures uniformity. Figure 2.6. shows the interpretation of length for a straight and a slightly deformed paprika. It is important to stress that the basis for measuring the length is not the imaginary centreline, but the actual length of the fruit. This is how the uniformity of the produce can be ensured in each packaging unit. Since heavily deformed pieces of produce are not required to be classified on the basis of size, their shoulder width is not necessary to be measured, either. Thus they are usually sorted on the basis of their length, if required (Figure 2.7).

## 2.6. Describing the Shape of the Paprika

It became obvious in the early phase of the research that the description of the paprika requires a unique data reduction system. The essence of the reduction algorithm is to determine the most important parameters of the given piece
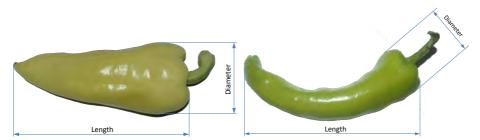
Figure 2.6 Interpreting the characteristic dimensions on sample pieces of paprika



Figure 2.7 For heavily deformed pieces of fruit, only the length of the fruit must be measured

of paprika by registering a small number of characteristics, and to ensure fast comparison of the measured values. The procedure elaborated is based on appointing the significant points of the paprika and determining the parameter that describes its curvature.

### 2.6.1. *Determining the Degree of Curvature of the Paprika*

Pieces of fruit with deformed shape must be recognized among the produce to be sorted by size, and such deformed pieces can only be classified as non-sorted or as special category ("lecsó", i. e. ragout). Their curvature, however, is to be determined, since there are types of paprika that are inherently "deformed". Beside considering the length, it is also necessary to distinguish between curved and deformed produce, therefore the accurate measuring of curvature is very important.

When the fruit's five basic points ($P_1[x_{P1}, y_{P1}]$, $P_2[x_{P2}, y_{P2}]$, $P_3[x_{P3}, y_{P3}]$, $P_4[x_{P4}, y_{P4}]$, $\overline{P_4}[x_{\overline{P4}}, y_{\overline{P4}}]$), are available, the solution developed by me will compute the curvature based on the relationship of these points and the center of gravity $S[x_S, y_S]$ (Figure 2.8). If we compute the angle $\gamma_1$ between segments $\overline{P_1 S}$ and $\overline{P_2 S}$ drawn on Figure 2.8 and 2.9, as well as angle $\gamma_2$ between segments $\overline{P_1 S}$ and $\overline{P_3 S}$, their ratio will provide an index number for the general curvature of the fruit, which can also be applied to determine the direction of the deformation.
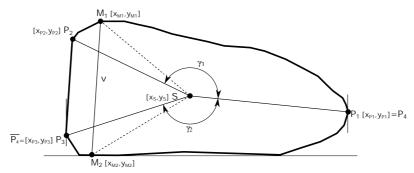
Figure 2.8 Cardinal points and segments appointed on a straight piece of paprika ($\overline{P_2P_3} \parallel \overline{M_1M_2}$)

Since at this point we leave the circle of cardinal numbers and accuracy requires the use of fractions in complex operations, we need to focus on solutions that can be computed at the fastest speed per unit. Where ever it was possible, I used the procedures optimized by the manufacturer of the microcontroller; where there was no such option, I used my own methods for the calculations. I also had to pay attention to values computed at a certain point of the process but also required in later operations, that they are calculated only once.

We calculate the length of segments $\overline{P_1S}$, $\overline{P_2S}$ and $\overline{P_3S}$:

$$\overline{P_1S} = \sqrt{(x_S - x_{P1})^2 + (y_S - y_{P1})^2}; \tag{2.1}$$

$$\overline{P_2S} = \sqrt{(x_S - x_{P2})^2 + (y_S - y_{P2})^2}; \tag{2.2}$$

$$\overline{P_3S} = \sqrt{(x_S - x_{P3})^2 + (y_S - y_{P3})^2}. \tag{2.3}$$

To determine the curvature, first we calculate the angle of inclination for segments $\overline{P_1S}$, $\overline{P_2S}$ and $\overline{P_3S}$:

$$\varphi\,\overline{P1S} = \begin{cases} \pi - \arccos \dfrac{x_S - x_{P1}}{\overline{P1S}}, & \textit{if } y_{P1} > y_S; \\[3mm] \pi + \arccos \dfrac{x_S - x_{P1}}{\overline{P1S}}, & \textit{if } y_{P1} < y_S; \end{cases} \tag{2.4}$$

$$\varphi\,\overline{P2S} = \begin{cases} \pi - \arccos \dfrac{x_S - x_{P2}}{\overline{P2S}}, & \textit{if } y_{P2} > y_S; \\[3mm] \pi + \arccos \dfrac{x_S - x_{P2}}{\overline{P2S}}, & \textit{if } y_{P2} < y_S; \end{cases} \tag{2.5}$$

$$\varphi\,\overline{P3S} = \begin{cases} \pi - \arccos \dfrac{x_S - x_{P3}}{\overline{P3S}}, & \textit{if } y_{P3} > y_S; \\[3mm] \pi + \arccos \dfrac{x_S - x_{P3}}{\overline{P3S}}, & \textit{if } y_{P3} < y_S. \end{cases} \tag{2.6}$$

Calculation of the angles $\gamma_1$ and $\gamma_2$:

$$\gamma_1 = \varphi\,\overline{P2S} - \varphi\,\overline{P1S}; \tag{2.7}$$

$$\gamma_2 = \varphi\,\overline{P3S} - \varphi\,\overline{P1S}. \tag{2.8}$$

Calculation of the supplementary angles $\gamma_1$ and $\gamma_2$:

$$\gamma_{(12)k} = 2\pi - |\varphi\,\overline{P3S}| - |\varphi\,\overline{P1S}|. \tag{2.9}$$

Finally, after $\gamma_1$ and $\gamma_2$ are determined, the factor describing the degree of curvature can also be calculated:

$$g = \begin{cases} \left|\dfrac{\gamma_1}{\gamma_2}\right|, & if \quad |\gamma_1| \geq |\gamma_2|; \\[2ex] \left|\dfrac{\gamma_2}{\gamma_1}\right|, & if \quad |\gamma_1| < |\gamma_2|. \end{cases} \tag{2.10}$$

For regular fruit shapes, the value of the $g$ "degree of curvature" is 1 – this is a ratio without any unit of measure, which increases in proportion with the increase of curvature. Due to intense computing needs, a future system may refine the calculation of the degree of deformation by fitting a circle on points $P_1$, $P_2$, $P_3$ , and comparing the areas above and below the paprika shape. If necessary, the two methods can also be used in combination, thereby increasing the reliability of the assessment.



Figure 2.9 Appointing the cardinal points and segments of a slightly deformed piece of paprika ($\overline{P_2P_3}\ \|\ \overline{M_1M_2}$)

### 2.6.2. *Determining the Dimensions and Parameters Characteristic of the Shape*

After the degree of curvature is calculated, the following parameters can be determined on the basis of the data registered earlier:

1. Length of the paprika (without stalk):

$$h = Max\bigl(|x_{P2} - x_{P4}|;\ |x_{P3} - x_{P4}|\bigr). \tag{2.11}$$

By length we mean the distance between points $P_2$ and $P_4$, or $P_3$ and $P_4$ on the $x$ axis. The length of the paprika without the stalk is the greater distance.

2. Total length of the paprika (with stalk):

$$h_t = \left| x_{P4} - \overline{x_{P4}} \right|. \tag{2.12}$$

The full length of the paprika is the distance between points $P_4$ and $\overline{P_4}$ that is the first and the last useful piece of information, measured on the $x$ axis.

3. The length of segment $\overline{P_2 P_3}$ , designating the baseline of the stalk, on the $x$ axis:

$$m = x_{P2} - x_{P3}. \tag{2.13}$$

The significance of determining this value is that in the next step, the position of points $M_1$ and $M_2$ are determined with the help of this distance.

4. The length of segment $\overline{M_1 M_2}$, that is the fruit's shoulder width, can be calculated with the help of iteration, using value m of the distance:

$$v = Max\left[ \sqrt{m^2 + (H_{K(i+m)} - H_{B(i)})^2} \right]. \quad (i\text{: only of } x_{P2} : x_{P3} \text{ environment}) \tag{2.14}$$

Search for the maximum width is conducted only in the immediate proximity of the stalk, whose position ($x_{P2} : x_{P3}$), had already been determined in earlier processes, and the direction of the search is appointed by the forward/backward ($EH$) factor. In order to reduce the amount of computing opeartions, the square root calculation can be omitted, since the modified shoulder width $\hat{v}$ will also do as a characteristic piece of information for the classification:

$$\hat{v} = Max\left[ m^2 + (H_{K(i+m)} - H_{B(i)})^2 \right]. \quad (i\text{: only of } x_{P2} : x_{P3} \text{ environment}) \tag{2.15}$$

In this case, the computing operation containing the square root calculation must be run only once after the search for the maximum width is finished, with the maximum value found during iteration, and thus a significant amount of runtime can be saved:

$$v = \sqrt{\hat{v}}. \tag{2.16}$$

5. Approximate length of the centreline:

$$h_{kk\ddot{o}z} = Max\left[ (\overline{P_1 S} + \overline{P_2 S}); (\overline{P_1 S} + \overline{P_3 S}) \right]. \tag{2.17}$$

By the approximate length of the centreline, we mean the total length of segments $\overline{P_1 S}$ and $\overline{P_2 S}$, or $\overline{P_1 S}$ and $\overline{P_3 S}$. By calculating this length, we can make conclusions concerning the correctness of the results obtained.

## 3.   RESULTS

In this chapter, I will set forth, analyse and assess the results of my research, as well as present the new scientific results. For that purpose, I will examine the appropriateness of the equipment and procedures developed for the operation of the visual shape recognition system, by which the images are taken and segmented, and the paprika's geometry and curvature is described and determined. After these, I will turn to the applicability of the developed method for other styles of produce, as well as to other options for generalization.

### 3.1.   Examining the Segmentation Procedures

Traditional classification procedures have been satisfactory for white and red coloured fruit and with clean conveyor belts, but for yellow (greenish yellow) and green fruit, the accuracy of appointing contours has been significantly reduced. The main reason for that is that relative to the lighting, these colours (yellow, greenish yellow and green) produce substantially lower contrast with the background than, for example, a white fruit.

Tests were conducted manually, and during these examinations I compared the actual data with those measured by the visual classification system.

I also defined the steps of checking the contour shots, as well as the rules for their assessment. Where it is not specifically given beside the measurement data, the measurements were repeated three times each and the Table contains the average of these measurements for each item.

Table 3.1 Results of contour shots, using proprietary classification based on identifying the contours

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Sweet, white | correct | 457 | 457 | 100% |
| | uncorrect | 0 | | 0% |
| Special, red | correct | 380 | 381 | 99,7% |
| | uncorrect | 1 | | 0,3% |
| Special, green | correct | 288 | 292 | 98,6% |
| | uncorrect | 4 | | 1,4% |
| Special, yellow/green | correct | 114 | 115 | 99,1% |
| | uncorrect | 1 | | 0,9% |
| Hot pepper, green | correct | 319 | 322 | 99,1% |
| | uncorrect | 3 | | 0,9% |

Since traditional procedures were unable to deliver an appropriately low rate of error for all fruit colours, I started looking for ways to identify contours with as low a rate of errors as possible. The algorithm developed for this purpose was expected to appropriately and accurately distinguish between the object and its background even under uncertain conditions.

From Table 3.1 it can be drawn that this proprietary contour detection method works very accurately even under poor lighting conditions and with produce of diverse colours that are unfavourable from the point of view of the lighting and that are greatly inhomogeneous geometrically, like special types and pointed, sharp types of paprika.

## 3.2. Examining the Elaborated Contour Detection Algorithm

In order to check the sturdiness of the contour detection algorithm, I examined the characteristics of its operation as well as the information it provided, under various extreme conditions. These included special circumstances that make the process of classification hard.

Table 3.2 Sensitivity of the contour detection procedure to the dirtiness of the conveyor

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Clean conveyor belt | correct | 128 | 129 | 99,2% |
| | uncorrect | 1 | | 0,8% |
| Moderately dirty conveyor belt | correct | 126 | 129 | 97,7% |
| | uncorrect | 3 | | 2,3% |
| Very dirty conveyor belt | correct | 124 | 129 | 96,1% |
| | uncorrect | 5 | | 3,9% |
| Clean, wet conveyor belt | correct | 129 | 129 | 100% |
| | uncorrect | 0 | | 0% |

Table 3.2. shows the effects of the conveyors dirtiness on the operation of the contour detection procedure. During the process I strived to include a wide variety of paprikas both in form and type. A clean conveyor belt meant the condition when the colour of the belt was closest to the lowest "black level", based on the measurements of the camera.

From the data of the Table it is obvious that it was impossible to create a production condition when the number of erroneous contour shots significantly increased. All measurements were repeated five times during the examination.

## 3.3. Determining the Geometrical Characteristics of the Paprika

The two main geometrical parameters used in the process of classification – which must also meet the requirement of consistency – are the resultive shoulder diameter $(V)$, and the resultive length $(H)$. It is important to note that length here is always meant without the stalk, and the shoulder width is the maximum width of the fruit.

The assessment of the stalk identification process was conducted in two phases. First, I examined the effects of the deformity on the accuracy of the algorithm, then the success rate of detecting a stalk that inclines back against the fruit. Table 3.3. contains the data to be examined during the procedure elaborated for defining the position of the stalk of deformed pieces of fruit. Based on measurement data it can be drawn that the error margin of the algorithm remained low, under the required 5% level, independent of the type of deformity.

Table 3.3 Success rate of finding the position of the stalk when it does not incline back against the fruit

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Straight harvest | correct | 213 | 216 | 98,6% |
| | uncorrect | 3 | | 1,4% |
| Hooked harvest | correct | 75 | 77 | 97,4% |
| | uncorrect | 2 | | 2,6% |
| Moderately hooked harvest | correct | 86 | 88 | 97,7% |
| | uncorrect | 2 | | 2,3% |
| Very hooked harvest | correct | 52 | 54 | 96,3% |
| | uncorrect | 2 | | 3,7% |
| Rumpled tip harvest | correct | 29 | 30 | 96,7% |
| | uncorrect | 1 | | 3,3% |
| Back hooked harvest | correct | 69 | 71 | 97,2% |
| | uncorrect | 2 | | 2,8% |

Based on data listed in Table 3.4, it is clear that if the stalk does not appear to be inclined, this condition can be detected and taken into account during the calculation of the length, and the proper operation of the algorithm is ensured. However, if the fruit is deformed, the detection of straight stalks was also less accurate, especially with heavily deformed pieces. But since classification based on size is not necessary or required for deformed pieces, in such cases a rough measurement of length is enough. As a next step, I examined the operation of the FAST corner detection algorithm fit in an embedded system and

Table 3.4 Success rate of finding the position of an inclining stalk

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Straight stem and paprika | correct | 40 | 40 | 100% |
| | uncorrect | 1 | | 0% |
| Stalk bends back (one view), straight paprika | correct | 32 | 33 | 97% |
| | uncorrect | 1 | | 3% |
| Stalk bends back (two view), straight paprika | correct | 19 | 25 | 76% |
| | uncorrect | 6 | | 24% |
| Straight stem, deformed paprika | correct | 33 | 35 | 94,3% |
| | uncorrect | 2 | | 5,7% |
| Stalk bends back (one view), deformed paprika | correct | 39 | 42 | 92,9% |
| | uncorrect | 3 | | 7,1% |
| Stalk bends back (two view), deformed paprika | correct | 14 | 20 | 70% |
| | uncorrect | 6 | | 30% |

using special parameters. During the operation, the algorithm was supposed to appoint the tip of the paprika.

Table 3.5. shows that the tip of the paprika can be determined with a 100% accuracy for straight pieces of fruit, whereas the error rate for curved, semi-curved and crumpled pieces remains under 5%. The more deformed the fruit is, the lower the reliability of tip detection is. This, however, is not significant, since such pieces of fruit need to be classified on the basis of their length only.

With respect to geometric features, the accuracy of both the shoulder diameter and the length was examined, with a margin $\pm 2$ mm for the length, and of $\pm 1$ mm for the shoulder width.

The accuracy data of the dimensions defined by the machine are shown in Table 3.6. Based on the results, it can be stated that for white and red types the likelihood of accurate measurements was 100%, whereas for other types the amount of inaccurately defined dimensions remained under the required 5% margin. Therefore it can be stated that the accuracy of major geometric dimensions is acceptable.

## 3.4. Examining the Shape Description and Determining the Curvature

The results of the examination concerning deformed pieces of fruit can be seen in Tables 3.7 and 3.8. Tests regarding deformed produce have been divided into two parts. First, I examined the hit rate (accuracy) of the algorithm in classifying produce with various shape specifications into categories corresponding their specific deformations. From the measurement data shown in

Table 3.7, it can be concluded that the error rate of classification never reached the 5% error margin.

As a second step, I examined the identification of deformed produce of various colours (Table 3.8), including the effect of the colour on the success of classification.

Table 3.5 Examination of the operation of the implemented FAST corner detector; the property to be identified was the tip of the paprika

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Straight paprika | correct | 188 | 189 | 100% |
| | uncorrect | 0 | | 0% |
| Hooked paprika | correct | 78 | 81 | 97,5% |
| | uncorrect | 2 | | 2,5% |
| Moderately hooked paprika | correct | 63 | 66 | 95,5% |
| | uncorrect | 3 | | 4,5% |
| Very hooked paprika | correct | 30 | 32 | 93,8% |
| | uncorrect | 2 | | 6,3% |
| Rumpled tip paprika | correct | 41 | 43 | 95,3% |
| | uncorrect | 2 | | 4,7% |
| Back hooked paprika | correct | 57 | 65 | 87,7% |
| | uncorrect | 8 | | 12,3% |

Table 3.6 Accuracy of machine defined length and shoulder width (shoulder diameter) dimensions

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Sweet, white | accurate | 142 | 142 | 100% |
| | inaccurate | 0 | | 0% |
| Special, red | accurate | 89 | 89 | 100% |
| | inaccurate | 0 | | 0% |
| Special, green | accurate | 44 | 45 | 97,8% |
| | inaccurate | 1 | | 2,2% |
| Special, yellow/green | accurate | 35 | 36 | 97,2% |
| | inaccurate | 1 | | 2,8% |
| Hot pepper, green | accurate | 71 | 73 | 97,3 % |
| | inaccurate | 2 | | 2,7% |

The greatest uncertainty was observed with green produce, but even in such cases the amount of faulty identifications remained under 5%. The identification of curved produce worked with a high hit rate, and errors were due primarily to pieces stuck together because of their deformed shape.

Table 3.7 Examining the automatic classification of the fruit based on the measured curvature

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Straight paprika | correct | 189 | 189 | 100% |
| | uncorrect | 0 | | 0% |
| Hooked paprika | correct | 79 | 81 | 97,5% |
| | uncorrect | 2 | | 2,5% |
| Moderately hooked paprika | correct | 64 | 66 | 97% |
| | uncorrect | 2 | | 3% |
| Very hooked paprika | correct | 31 | 32 | 96,9% |
| | uncorrect | 1 | | 3,1% |
| Rumpled tip paprika | correct | 42 | 43 | 97,7% |
| | uncorrect | 1 | | 2,3% |
| Back hooked paprika | correct | 62 | 65 | 95,4% |
| | uncorrect | 3 | | 4,6% |

Table 3.8 The influence of the fruit's colour on determining the extent of deformation

| Description | Rating | Number of cases [pcs] | Total cases [pcs] | Ratio of cases [%] |
|---|---|---|---|---|
| Sweet, white | correct | 121 | 123 | 98,4% |
| | uncorrect | 2 | | 1,6% |
| Special, red | correct | 116 | 118 | 98,3% |
| | uncorrect | 2 | | 1,7% |
| Special, green | correct | 109 | 114 | 95,6% |
| | uncorrect | 5 | | 4,4% |
| Special, yellow/green | correct | 37 | 39 | 97,4% |
| | uncorrect | 1 | | 2,6% |
| Hot pepper, green | correct | 98 | 103 | 95,8% |
| | uncorrect | 5 | | 4,9% |

## 3.5.    Examining the Manual and the Automated Classification

In conducting the tests comparing manual and machine sorting, I selected the class of produce between 50-59 mm, and the task was to manually select pieces within this specific range. The basis for classification was the shoulder diameter for both the manual and the machine sorting. Pieces of fruit selected manually were then measured and assigned a numeric identifier, before being reloaded to the classifying machine. The procedure was repeated five times. Length was checked with the help of a caliper, whereas diameter was determined with a punch plate screen during manual classification.

Manual classification was conducted by sorting, which means that the workers had to pick pieces corresponding to the required size. The tests were run on two samples, normal (sample marked $N$) and careful (sample marked $G$), referring to the specific sorting conditions. In 'normal' conditions, the personnel did not know that their work would be assessed by measurement, whereas in 'careful' conditions they were informed of this circumstance beforehand. Related measurement results can be seen in Tables 3.1-3.4. and 3.9.

From the data of the manual sorting it can be stated that under 'normal' conditions (Diagram 3.1), sorting discipline was lower than under 'careful' circumstances (Diagram 3.3). From these diagrams it can also be concluded that during manual sorting there was a tendency to uprate the produce, that is many pieces were classified into the specified category even though they were smaller.

Therefore it is important to note that the frequency shown by the histogram provides only indirect information as to the accuracy of sorting, since the number of pieces classified into a given category is influenced by the geometric characteristics of the produce and their distribution. Therefore during the comparison and assessment of manual and machine sorting, only such characteristics can be taken into account that are clearly identifiable on the basis of the histogram. Data in Table 3.9 compiled on the basis of diagrams 3.1-3.4 show that the error margin of machine sorting in both cases was around 3%, which is significantly lower than the 54% margin of 'normal' manual sorting, as well as the 37% margin of the 'careful' manual sorting.

Table 3.9 Numeric comparison of manual and machine sorting, with a shoulder diameter range of 50-59 mm

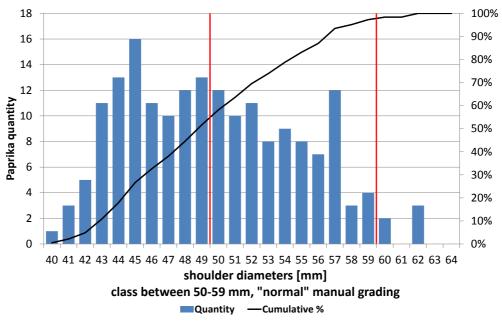| Sorting method | Min. [mm] | Max. [mm] | Scope [mm] | Avg. [mm] | Dev. [mm] | Sum. [pcs] | Fault [pcs] | Error rate [mm] | Error rate [%] |
|---|---|---|---|---|---|---|---|---|---|
| Manual, sample $N$ | 40 | 62 | 20 | 49,7 | 5,13 | 184 | 100 | -10; +3 | 54,3% |
| Manual, sample $G$ | 45 | 64 | 19 | 50,6 | 4,32 | 123 | 45 | -5; +5 | 36,5% |
| Mach., sample $N$ | 50 | 60 | 10 | 54,3 | 2,67 | 94 | 3 | -0; +1 | 3,2% |
| Mach., sample $G$ | 49 | 60 | 11 | 53,9 | 2,72 | 92 | 3 | -1; +1 | 3,3% |

Figure 3.1 Histograms of 'normal' manual sorting (sample $N$), the two vertical red lines representing the extremes of the size category
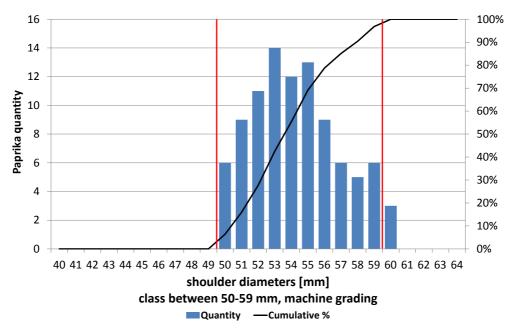


Figure 3.2 Histograms of the machine sorting of the 'normal' ($N$) sample, the two vertical red lines representing the extremes of the size category
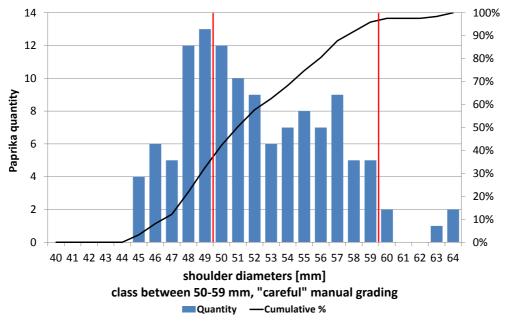
Figure 3.3 Histograms of 'careful' manual sorting (sample $G$), the two vertical red lines representing the extremes of the size category
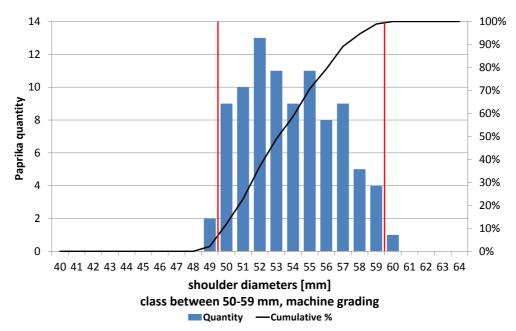


Figure 3.4 Histograms of the machine sorting of the 'careful' ($G$) sample, the two vertical red lines representing the extremes of the size category

### 3.6. Applicability of the Elaborated Method for the Classification of Other Types of Produce

Based on the favourable results of the developed classification procedures, I conducted preliminary experiments with other types of oblong vegetables that are rather insensitive to impact.

The measurement algorithm of the paprika was partly redesigned for this purpose, since the algorithms used for identifying the stem and the geometry of the paprika cannot be used for other types of produce. As a result, the parameters measured were the diameter, the length and the curvature.

During the tests, I strove to comply with the limit parameters of the shapes measured by the machine. Based on the measurement result, it can be stated that the contour defining procedure is capable of identifying the contours of other types of produce with a fairly low error margin, but because of the algorithm specifically developed for the paprika, this requires not only software modifications but also constructional ones.

Measurement results related to the machine's universal operation are shown in Table 3.10. Based on these data received, it can be stated that, with minimal changes, the error margin of the elaborated procedure always remains under the required 5%. During preliminary tests it also turned out that this result can be further reduced by optimizing the procedures.

Table 3.10 Examining oblong produce with the extended and redesigned algorithms; during the examination, the success of the contour shot was checked, and the two geometrical features (width and length) as well as the curvature were measured

| Description | Rating | Number of cases [*pcs*] | Total cases [*pcs*] | Ratio of cases [%] |
|---|---|---|---|---|
| Potato | correct | 392 | 409 | 95,8% |
| | uncorrect | 17 | | 4,2% |
| Long cucumber | correct | 32 | 33 | 97,0% |
| | uncorrect | 1 | | 3,0% |
| Picking cucumber | correct | 89 | 91 | 97,8% |
| | uncorrect | 2 | | 2,2% |
| Carrots (without leaves) | correct | 30 | 30 | 100% |
| | uncorrect | 0 | | 0% |
| Zucchini | correct | 10 | 10 | 100% |
| | uncorrect | 0 | | 0% |

## 4. NEW SCIENTIFIC RESULTS

The scientific results of my research can be summed up by topics as follows:

1. *Elaboration of a segmenting method for determining the contour of the paprika*
I have created an innovative, complex algorithm that designates the contours of the paprika by analysing the derivative of the image section function, as opposed to earlier procedures applying thresholding (comparing). With the method created, it possible to reduce the number of faulty contour shots by up to 98%, even under sub-optimal production conditions. Under optimal conditions, on the other hand, that is with a clean and light reflecting surface, the error margin of the procedure is less than 1,5%.

2. *Procedure for determining the shape parameters of the paprika*
I have created a method to define the most characteristic geometric points of produce, the so-called base points, which describe the morphological features of each individual piece in a manner that allows for the classification to take place accurately even at high speed. The limit value parameters were assigned to each specific size group. I have also proven that this theoretically and functionally new and modern system is more effective and more accurate than the manual methods even at high rates of productivity.
The complex theoretical assessment system was validated by being integrated into a system capable of on-site classification, through which I have identified the sorting errors that may occur in practice. In the developed system, the amount of errors remained under the 5% rate corresponding to requirements, therefore machine classification is realized with an average shoulder diameter variation of 1,0-1,5 mm, as opposed to manual sorting, where the variation is $\pm$5-10 mm.

3. *Ratio for describing the deformation of conic and pointed types of paprika*
I have introduced a new morphological factor, the degree of curvature ($g$) of conic and pointed types of paprika, calculated on the basis of the base points and the center of gravity of the shape. As part of that, I also defined the major groups of the typical deformations of the paprika. The ratio elaborated for providing the curvature of the paprika, without dimension, is the following:

$$g = \begin{cases} \left|\dfrac{\gamma_1}{\gamma_2}\right|, & \text{if} \quad |\gamma_1| \geq |\gamma_2|; \\[3ex] \left|\dfrac{\gamma_2}{\gamma_1}\right|, & \text{if} \quad |\gamma_1| < |\gamma_2|. \end{cases} \qquad [-]$$

4. *Generalization of the shape description and contour detection procedure*

   I have established that with the help of the equipment elaborated theoretically and then set up based on this theory, a contour detection and shape description algorithm can be generalized based on the analysis of the image segments, and this algorithm can be applied to other types of fruits and vegetables that are less sensitive to impact (such as cucumbers, carrots, asparagus, potatoes, celery, onions etc.) for the purpose of their sorting and classification. With the modified method, the error margin always remained below the 5% rate during the tests. The software's parameters need to be adapted to the specific type of produce.

## 5.   CONCLUSIONS AND SUGGESTIONS

With the help of the procedures elaborated during my research I created the method capable of automatically classifying conic and pointed types of culinary paprika. I have also developed an autonomous embedded system for the identification of processes and models, as well as for the examination of algorithms, and as a result, sorting has become easy to do, even under difficult, on-site conditions, at low cost and with high accuracy.

The elaborated method allows for the detection of the contours of various types of paprika with varying shapes and colours, as well as for deducing the stem from the contour data, for determining the characteristic points of the paprika, for the description of the curvature and for determining the major dimensions, thereby allowing for the description of the shape. According to present practice, culinary paprika is mostly sorted manually, which cannot exclude human subjectivity. This means that paprika usually comes to the market in sizes that are characteristic of the producer and their personnel, rather than of the produce itself. This, in turn, leads to significant disadvantages for certain producers when selling their products. Machine sorting has several advantages compared to manual sorting:

- Greater accuracy and uniformity,
- Objectivity,
- Speed.

As a result of the development, a segmentation method was elaborated, which, after proper redesigns, may be applicable in other areas as well. It can be stated that its application is not necessarily narrowed down to detecting the contours of agricultural produce, but may also work in an industrial environment where the same disadvantageous conditions apply to the imaging process, and the measurements must be conducted in real time and at low cost.

The application of methods elaborated for defining the size, shape and curvature is more limited, though, since when other types of agricultural produce are examined, the procedure must by all means be specialized to the characteristics of the given type of produce (an onion, for example, or a cucumber, requires totally different methods to interpret their geometry or curvature).

During the research, the algorithm elaborated for paprika was partially reprogrammed to be capable of sorting other types of oblong produce. Based on the work done so far, it can be stated that the geometric characteristics of oblong types of produce, such as cucumbers, carrots, zucchini, asparagus, potatoes, can be determined with relatively few modifications.

# 6.   SUMMARY

## INNOVATIVE CLASSIFICATION OF THE CULINARY PAPRIKA WITH AN AUTOMATIC IMAGING SYSTEM

The purpose of my research was to find a solution for the classification of conic and pointed types of culinary paprika, with which such classification can be done accurately, rapidly and observing all standard requirements.

After elaborating the theoretical bases of automated sorting, I developed a procedure and method that can be used in industrial circumstances for identifying the contours of culinary paprika at high speed, which in turn can be used to sort the produce at high speed. When looking for the right method, I found that using a line CCD sensor is the most appropriate solution for taking high-speed shots of the contour. In order to filter out curved pieces, I elaborated a procedure for describing the shape of the paprika. This procedure is capable of describing the major characteristic features of the fruit with the help of a few parameters, but it was also established that the degree of curvature can be described even with an auxiliary parameter.

After the mathematical formulation of the procedures, the necessary process identification was conducted with the help of an actual, constructed sorting machine. The sorting process needed for the identification was run in a modern, embedded measuring and processing system capable of real-time operation. I matched the elaborated procedures to the actual sorting process, and then tested their accuracy and features under diverse operational circumstances. Based on the results it was established that micro-controller based shape recognition seems to be a feasible method for other types of uses as well.

The scope of using the generalized procedure can be rather wide, since there is great demand for robust and fast segmentation methods. As for the segmentation procedure, its application is not necessarily narrowed down to detecting the contours of agricultural produce, but may also work in an industrial environment where the same disadvantageous conditions apply to the imaging process that have been discussed in this paper. Procedures for describing shape and curvature can be generalized with limitations, since every type of shape requires its own parameters to be applied to the elaborated method. In this context, I examined the applicability of the algorithm for certain oblong types of vegetables that are rather insensitive to impact. Based on the results and after the construction of the prototype, several systems were built with approximately the same technical content, and the assessment of their operation is ongoing.

## 7. MOST IMPORTANT PUBLICATIONS RELATED TO THE THESIS

*Referred articles in foreign languages*

1. **Gergely, Z.** – Judák, E. (2008): A microcontroller based algorithm for sorting white paprika. Hungarian Agricultural Engineering, No. 2007-20. pp. 37-39. HU ISSN: 0864-7410
2. **Gergely, Z.** – Beke, J. (2013): Morphological algorithm for fast contour characterization in white paprika sorting process. Mechanical Engineering Letters, Release of Trans-Trio Sciences, vol. 9/2013. pp. 98-103. HU ISSN: 2060-3789
3. **Gergely, Z.** – Beke, J. (2013): A micro-controller-based algorithm for fast and robust edge detection in white paprika sorting process. Mechanical Engineering Letters, vol. 10/2013. pp. 161-169. HU ISSN: 2060-3789
4. **Gergely, Z.** – Petróczki, K. – Beke, J. (2016): A high performance method for sorting white paprika. The Experiment Journal, 2016. ISSN: 2319-2119 (IF: 0,33) (accepted, in press)

*Referred articles in Hungarian language*

5. **Gergely, Z.** – Judák, E. (2008): Automatizált paprikaválogatás beágyazott alakfelismerő rendszerrel. Mezőgazdasági Technika, XLIX. évf. 2008/11. 2-4. o. HU ISSN: 0026-1890
6. **Gergely, Z.** – Beke, J. (2015): Az osztályozási hibák csökkentésének lehetőségei a HPV-I sorozatú paprikaválogató gépeken. Mezőgazdasági Technika, LVI. évf. 2015/11. 2-4. o. HU ISSN: 0026-1890
7. **Gergely, Z.** – Beke, J. (2016): A kúpos és hegyes alakkörbe tartozó étkezési paprikatermések alakjellemzése. Mezőgazdasági Technika, LVII. évf. 2016/7. 2-5. o. HU ISSN: 0026-1890