From real to virtual eyes: a classification almost 4.0 tomatoes

Mariana Matulovic, Cleber Alexandre de Amorim, Angela Vacaro de Souza, Paulo Sérgio Barbosa dos Santos, Geovane Yuji Aparecido Sakata, Guilherme Pulizzi Costa, Douglas Cardozo de Almeida, Jéssica Marques de Mello

Abstract

The change in the color of the vegetables peel during the ripening process is the main criterion used by the consumer to define the fruit ripeness degree and for the producer to determine the best time of harvest. This relationship between bark coloration and different maturation stages allows the producer to establish harvest planning and extend shelf life. Students and faculty of the Biosystems Engineering course at São Paulo State University (UNESP), Tupã Campus, designed and developed a low-cost prototype of a fruit sorting belt, specifically for cherry group tomatoes. In the future, improvement in machinery with the insertion of new devices such as cameras, embedded system, combines sensor technology 3.0 with machine learning 4.0.

Keywords: Agricultural machinery, sensing, technologies 3.0, CIELAB.

1. Introduction

The change in the color of the vegetables peel during the ripening process is the main criterion used by the consumer to define the fruit ripeness degree and for the producer to determine the best time of harvest. This relationship between bark coloration and different maturation stages allows the producer to establish harvest planning and extend shelf life.

Tomato is one of the most consumed vegetables in the world, with a production of over 182 million tons [7]. In Brazil, production was approximately 4.5 million tons in 2018, is the second most-produced vegetable [20]. In Brazil, the normative instruction No. 009 of 2002 says about the selection and classification of fruits and vegetables [1]. This one is a process that occurs according to size and shape, color, maturity point, and defects. What's more, a high-quality product will have a higher added value. Better quality and reduced product loss can be achieved through technology applied to product selection and classification. This process may be less subjective; the more direct contact is avoided. Thus, a process performed by machine vision and later for processing and analysis of images will have a closer response to a pattern among the selected products.

Stinco et. al. used tomato coloration to measure and quantify changes in chemical composition and physical attributes of fruits [4]. Thus, computer vision has been used to measure the color of different foods objectively. These provide an advantage when compared with the conventional colorimeter, such as the ability to analyze each pixel of the entire surface, quantifying surface characteristics, and defects [19; 3].

The application of sensors, processing techniques, and image analysis has been used to obtain qualitative and quantitative data regarding fruit selection and classification. Process automation allows standardization at speeds higher than those achieved by manual and/or mechanical processes.

2. Theory

2.1 The tomato

Different types of tomatoes are available on the market: Carmem, persimmon, Dutch, Italian, and cherry. Each for a different purpose and found both fresh and processed (sauce). In the vast majority of supermarkets, they sell these fruits in bulk and not in packaging that could increase their shelf time. This context leads to a devaluation of products and often leads to increased loss rates within the production chain [2].

Tomato is a fleshy and juicy berry with a different appearance, size, and weight, depending on the cultivar. Most fruits are bright red when ripe [6]. The choice of tomato fruit was due to maintaining their active metabolic activities in the post-harvest period. At this stage, physiological and biochemical transformations occur, causing physicochemical changes in the fruit [17].

According to Andreuccetti et al. [2], the absence of measurable quality standards and packaging improvement maximizes problems in the production chain. To ensure tomato supply throughout the year, wholesalers receive the product from various regions of the country under different conditions and times of the year. This diversity of product origin makes it difficult to standardize. However, current market demands such as better quality, appearance, texture, taste and aroma, nutritional value, and food safety make producers feel the need to implement standards for classification. The broad acceptance of the vegetables is due to their organoleptic characteristics. The color one of the leading and decisive at the time of purchase [5].

2.2 CIELAB and data clustering algorithm K-Means

Clustering data to describe them according to their homogeneity, characteristics, or organizing them into a set of categories is known as clusters. There are several techniques used for data grouping, and a measure of distance, correlation, or association must be defined so that it is possible to determine (dis)similarity between objects. Such (dis)similarity can be calculated in countless ways, but each type of measure addresses a perspective, and its application depends on the expected objective and the kind of data worked. It is called here a cluster (or a partition) is a set of groups [20].

Data grouping algorithms are based on obtaining k groups of similar objects according to some pre-established criteria. However, the value of k must be provided by the user, informing an inappropriate amount can create problems of interpretation of the partition obtained. One of the primary data clustering algorithms is K-means. Initially sets up how many clusters are being sought; this is the parameter k. Thus, k points are randomly chosen as cluster centers. The Euclidean distance metric is the distance of all elements belonging to the group to parameter k. Then, the average distance in each cluster is calculated and is defined as new core values for their respective clusters. The whole process is repeated with new cluster centers [10].

The iteration continues until the elements contained in each cluster no longer change. After the iteration stabilizes, each point is assigned to the nearest cluster center. They are minimizing the total squared distance of all points to their cluster centers. The k-means algorithm produces excellent results when there are no overlapping groups.

Pattern recognition in images goes through a process of translating the image into numerical values (Several Color Coordinate Systems can describe the color of an Object). RGB (red, green, and blue concentration) is the best-known system, accompanied by the CIE L* a* b* (Commission Internationale de l'Eclairage's) [15]. Colors are represented in the CIELAB system described by the parameters L*, a*, and b* or the use of cylindrical luminosity (L*), hue (H*), and chroma (C*) coordinates that are related to the Munsell coordinate. The parameter a* informs about the development of the red color of the fruit and its degree of ripeness [13]. Positive values of a* indicate red color and negative to green color. The b * coordinate refers to yellow if the value is positive and blue when negative. L* indicates lightness (black-white) [11]. The H* parameter indicates the color quality, referring to the hue. Chroma is the quality of color that allows contrast to be distinguished by relating to the amount of color that exists. The chroma is associated with saturation, intensity, vibrancy, purity, chromaticity, and depth [14].

L* and H* values decrease proportionally to fruit ripening. This decrease indicates browning of fruit due to changes in the level of pigment, change the green to red color, for example. Concomitantly with this process, the value of a* and C* increase with fruit ripening [4, 21]. Tomatoes are always classified when they are ripe, divided by the color group, which can be red, pink, orange, or yellow. This subgroup classification analyzes tomato ripening and determines the color change of its peel. However, it is a visual and subjective classification and given the number of variations found in this type of fruit. Further analysis is required both for determining harvest timing and producer compensation, as well as for follow-up on the production line and post-harvest life. Therefore, it is necessary to promote the application of new technologies for automated classification of vegetables by color so that there is any contact between the machine, product, and labor.

Image processing, which consists of capturing and processing them, of improving visual information for human interpretation, or improve it for automatic machine perception/interpretation. Another area is computer vision, which consists of extracting information from images, such as locating objects and identifying changes in the environment, making the robot "see" the work environment [18].

3. Material and Methods.

3.1 Fruit Sorting Conveyor Belt

Figure 1 shows the layout of the machinery and instruments and devices that have integrated the system. Figure 1a presents the prototype designed in Inventor Professional 2017 software, Autodesk version Build 142.

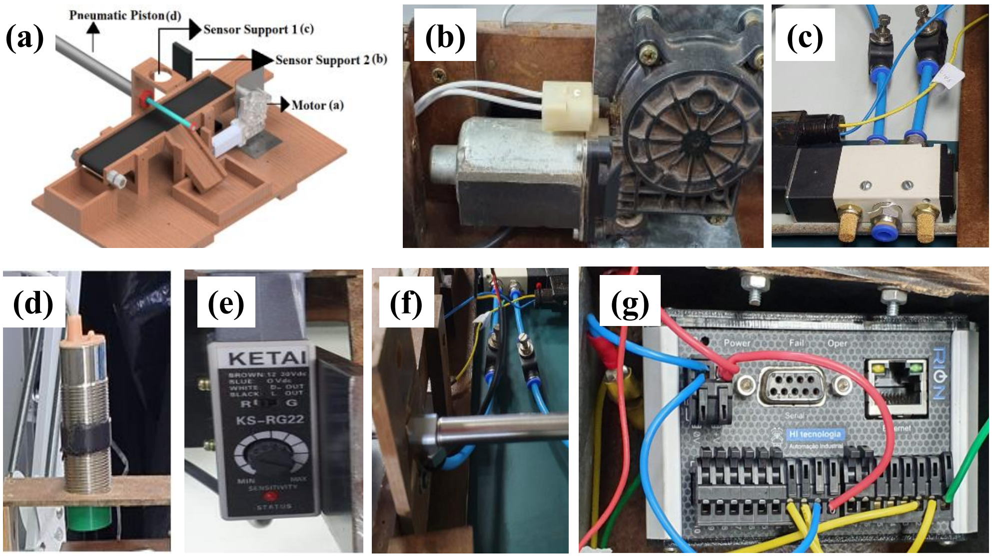

Figure 1. Prototypes and instruments used for the development of the fruit sorting belt: (a) Overview of the belt structure; (b) Bosch engine; (c) Color Sensor KS-W22; (d) Sensorbras Capacitive Sensor; (e) Air Cylinder CJ2; (f) Solenoid valves and (g) Programmable Logic Controller.

1. Simulations and theoretical calculations were performed, and the following set of sensors and actuators were defined for the prototype manufactured:
2. Bosch 12V electric motor. The device was salvaged from an automobile windshield cleaner, Fig. 1b;
3. High sensitivity color sensor, precision, and response time in the range 0.1-0.5 ms, KS-W22 KS-G22 KS-R22. This device detects the red and green colors, Fig. 1c;
4. Capacitive Proximity Sensor, 5 mm sensory distance, sensitivity adjustment, supply voltage 10 to 30 Vdc, Sensorbras - PNP, Fig. 1d;
5. Single-acting double-acting double-acting cylinder, 150 mm rod stroke, 16 mm bore, CDJ2B series, as shown in Fig. 1e.
6. Válula solenoid with two-position five modes, supply voltage 12 Vcc, inputs, and outputs 1/4 inch, Fig. 1f;
7. Relay Programmable remote I/O HI Technology, Rion model (HIO115), supply voltage 10 to 30 Vdc, eight digital inputs, four digital outputs, three analog inputs, one encoder two fast counters and one PWM communication Ethernet or serial cable, as shown in Fig. 1g.

Additionally, a treadmill control was inserted with two commands: (i) to start the engine rotation and (ii) to stop the motor movement. The color sensor was allocated to a fixed support on the side of the conveyor, perpendicular to the direction of its; another support served to the capacitive coupling sensor, so as to leave the face perpendicular to the conveyor band; In this same support, a pneumatic piston was placed perpendicular to the direction of movement of its. The final conveyor is shown in Fig. 2a-b.

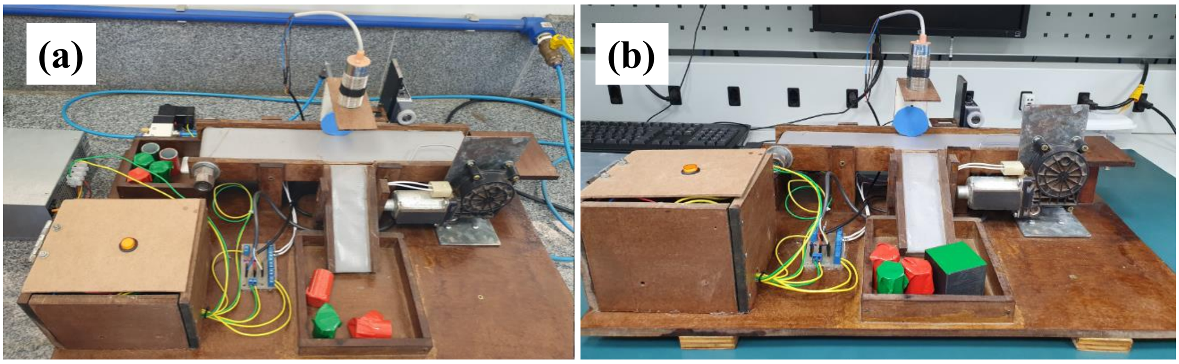

Figure 2. Final conveyor, top (a) and (b) front view.

After the equipment was built, the device was programmed through the programmable relay and remote I/O to automate the process. The developed program for the control conveyor consists of three steps:

1. Switching On/Off the conveyor motor via a holding button, i.e., the conveyor motor is On while the button is engaged;
2. Object color recognition. If the color chosen by the developer is green, then the answer is stored as a value of 1; otherwise, a value of 0 is stored.
3. Verification and implementation of the following procedure:
    a) Object detection by Capacitive sensor. When an object is detected, it checks what value was stored by the color sensor;
    b) If the color is red, stored the value 0, the pneumatic system starts to act, separating the object from the others;
    c) Otherwise, green, the object continues its path until it reaches the container at the end of the conveyor.

The procedure described above was carried out using a ladder programming language, as described below. First created a button (I0) to turn On/Off the conveyor selector (O0). The input I1 has addressed the detection of the color sensor, and to assist the storage of the results of this sensor was used R0 addressing. After measuring the color sensor, the code analyzes the detection of the object by the photoelectric sensor, addressed by I2. If the photoelectric sensor and auxiliary R0 have a high level, the auxiliary variable R2 is left in Set mode, and R0 is reset. This will trigger the piston, and the R0 will be ready for a new color sensor detection. The piece of code representing such functions is shown in Fig 3a.

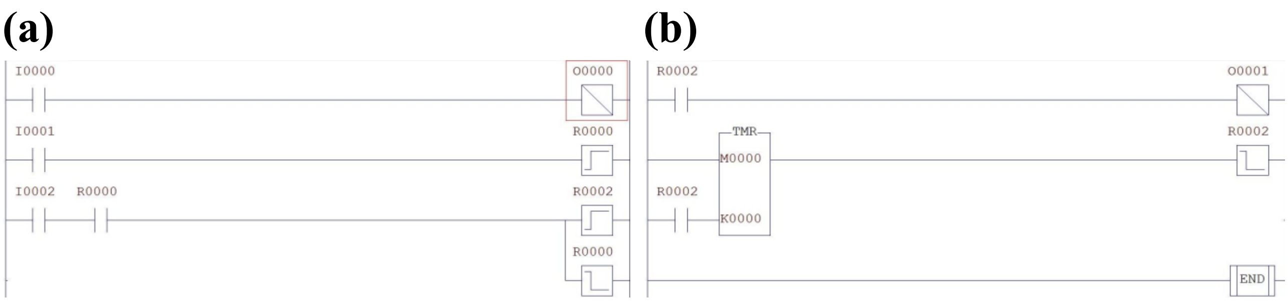

Figure 3. Programming code that control the conveyor, (a) First and (b) second part.

Finally, it can be stated that only when both sensors detect the object will the piston (O1) be driven, pushing the object off the belt. To monitor the response time between piston activation and movement, we have inserted a timer into the code. Code is presented in the Fig. 3b. Due to the distance between the color sensor and the pneumatic piston, it is not necessary to use a method if more than one object is detected before separation occurs. At this point, the actuator has already been triggered, even when inserted objects close together.

3.2 Recognition of red and green patterns

In order to differentiate the green fruit of red, it started the description of a standard image process, as seen in Fig. 4.

Figure 3. Standard imagem processo of Cherry Tomato

Due to the standard image composition, the CIELAB descriptor was chosen using 4-pixel sub-images. The result is a matrix with the color data of each of the pixels. Thus, using the k-means algorithm grouper, I have been able to image segmentation by separating into three different groups.

4. Results, Conclusions and Open Problems

The prototype calibration was done with different objects of red and green colors. A signal conditioning system is coupled to the device to circumvent the problem of the color sensor has a short-range. The piston speed, both in expansion and retraction, has been calibrated. Also, it has made adjustments in the photoelectric sensor so that the conveyor canvas did not interfere with the capture of data.

After the initial tests, they were replaced by cherry tomatoes. The results obtained were positive, presenting many imperfections due to several factors, such as close tones, inhomogeneity of color, rounded shapes, causing fruit-bearing and problems in its detection by color and photoelectric sensors. Next, there is a photo sequence of positive results for the selection of tomatoes.

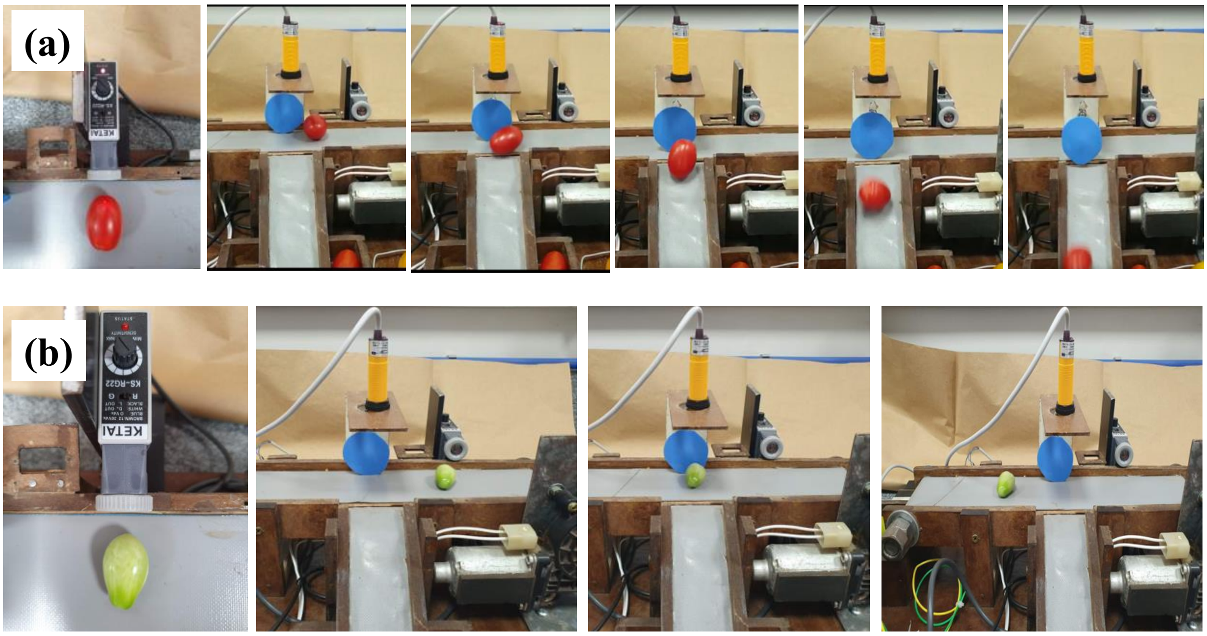

Figure 4. Sequence of the fruit selection process: (a) tomato red color detection, Photoelectric sensor detection, Start of piston drive, Push piston initial and End of ripe tomato separation process; (b) tomato green color detection and Photoelectric sensor detection and process continuity

In order to alleviate, and even remedy, the execution problems presented in the belt sensor set, we link to the mature green fruit recognition system through a pattern recognition algorithm, namely: K-means. The images were matrixed using the CIELAB, generating a database that was clustered by some predefined mathematical algorithm - such as Euclidean distance measurement. The results obtained were satisfactory as shown in Fig. 6a-b.

Figure 5. (a) Green pattern and (b) red pattern recognition.

The linking of the developed algorithm with the conveyor programming is in the execution phase. In a next step the programmable logic controller will be changed by the microcontroller MSP430FR5949. It is hoped that optimize and lower the machinery, in addition to embark developed artificial intelligence; inserting a model VGA 0V7670 camera into the conveyor so that sensor technology 3.0 brings us great results when added to machine learning 4.0.

6. Acknowledgement

This work was supported by Brazilian National Council for Scientific and Technological Development - CNPq [grant numbers: 421782/2016-1] and São Paulo Research Foundation - FAPESP [grant numbers: 2019/00021-4].

7. References

[1] Brasil, “Instrução Normativa no 009, de 12 de novembro de 2002. Embalagens de produtos hortícolas”, Diário Oficial, Brasília, 12 nov, 2002.

[2] C. Andreuccetti, M. D. Ferreira, A. S. D. Gutierrez, M. Tavares, “Classificação e padronização dos tomates cv. Carmem e Débora dentro da CEAGESP – SP”, Jaboticabal, v.24, n.3, 2004. pp.790-798.

[3] C. J. Du, Sun, D. W, “Recent developments in the applications of image processing techniques for food quality evaluation”, Trends in Food Science & Technology, v. 15, n.5, 2004, pp.230 -249.

[4] C. M. Stinco, F. J. Rodríguez-Pulido, M. L. Escudero-Gilete, et al., “Lycopene isomers in fresh and processed tomato products: correlations with instrumental color measurements by digital image analysis and spectroradiometry”, Food Research International, Ottawa, v.50, n.1, 2013, pp.111-120.

[5] Companhia de entrepostos e armazéns gerais de São Paulo, “Centro de Qualidade em Horticultura”, Programa Brasileiro para Modernização da Horticultura: Normas de classificação de tomates, São Paulo, 2003.

[6] F. A. R. Filgueira, “Novo manual de Olericultura: agrotecnologia moderna na produção e comercialização de hortaliças”, Viçosa-MG: UFV, 2008.

[7] Fao, “Food and Agriculture Organization of the United Nations”, Disponível em: < http://www.fao.org/statistics/en/>, Acesso em: out.2019

[8] G. Piatetsky-Shapiro, et al., “Advances in knowledge discovery and data mining”, Menlo Park: AAAI press, 1996.

[9] Globo rural, “Produção de tomate será 1,2 % inferior à prevista em janeiro”, disponível em <https://revistagloborural.globo.com/Noticias/Agricultura/Hortifruti/noticia/2018/03/producao-de-tomate-sera-12-inferior-prevista-em-janeiro.html>, Acesso em: mar.2018

[10] I. H. Witten, et al., “Data Mining: Practical machine learning tools and techniques”, Morgan Kaufmann, 2016.

[11] J. Carneiro, "Análise da reflectância de argamassas", Relatório Técnico, Braga: Universidade do Minho, 2010.

[12] K. L.Yam, S. E. Papadakis, "A simple digital imaging method for measuring and analyzing color of food surfaces", Journal of Food Engineering, v. 61, n.1, 2004, pp. 137-142.

[13] L. Ahmed, A. B. Martin-Diana, RICO, D. Rico, et al., "Quality and nutritional status of fresh-cut tomato as affected by spraying of delactosed whey permeate compared to industrial washing treatment", New York, v. 5, n. 8, 2011, pp. 1-12.

[14] L. M. Régula, "Padrões virtuais e tolerâncias colorimétricas no controle instrumental das cores", Dissertação (Mestrado em Metrologia para a Qualidade Industrial)- Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

[15] P. B. Pathare, U. L., F. A. AL-Said, "Colour measurement and analysis in fresh and processed foods: a review", Food and bioprocess technology, v. 6, n. 1, 2013, pp. 36-60.

[16] R. D. Tillett, "Image analysis for agricultural processes: a review of potential opportunities", Journal of agricultural Engineering research, v. 50, 1991, pp. 247-258.

[17] R. Perveen, et al., "Tomato (Solanum Lycopersicum) carotenoids and lycopenes chemistry; metabolism, absorption, nutrition, and allied health claims-A comprehensive review", Critical Reviews in Food Science and Nutrition, 55 (7), 2015, pp. 919-929

[18] R. S. Castro, J. M. O. Barth, J. V. Flores, A. T. Salton, "Modelagem e implementação de um sistema ballandplate controlado por servo-visão", XI Simpósio Brasileiro de Automação Inteligente, Fortaleza, 2013.

[19] T. Brosnan, D. W. Sun, "Improving quality inspection of food products by computer vision—a review", Journal of Food Engineering, v. 61, n. 1, 2004, pp. 3-16.

[20] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "From data mining to knowledge discovery in databases", AI Magazine, v. 17, n. 3, 1996, pp. 37-37.

[21] Z. Zhang, L. Liu, M. Zhang, et al., "Effect of carbon dioxide enrichment on healthpromoting compounds and organoleptic properties of tomato fruits grown in greenhouse", Food Chemistry, Barking, v.153, 2014, pp.157-163.