Optimization of Brazil-Nuts Classification Process through Automation using Colour Spaces in Computer Vision

Sonia Castelo-Quispe 1, July Diana Banda-Tapia 1, Mónica N. López-Paredes 1, Dennis Barrios-Aranibar 1 and Raquel Patiño-Escarcina 1

1 Research and Software Development Center - Cathedra CONCyTEC in TIC’s of the Saint Agustín National University Arequipa, Perú
{scastelo2, inf.jdtapia, kaloni25, dennisbarrios, raquel.patino}@gmail.com

Abstract: The Brazil-nuts classification is a process where the brazil-nuts that do not present damages are classified according to its size: Large, Medium, Small and Tiny, prior to its exportation. The current method used for this process is a manual one, presenting several deficiencies, because it is subjective, slow and imprecise.

In this sense, this study proposes the application of computational vision to automate this process, considering that there is a direct relation between the weight and size of a Brazil-nut, using a conversion factor and the Brazil-Nut’s areas in order to estimate the weights and infer the type. The segmentation was obtained using the YCrCb colour space with a dynamic threshold for binarization, since the background of the images change by external factors such as illumination. The experimental results show that the performance achieved by this approach is 99.7%.

Keywords: Brazil-nuts classification, Colour spaces, Estimation of weights by area, Automation.

I. Introduction

Nowadays, one of the main economic activities in cities near the forest such as Madre de Dios in Peru, the north of Bolivia, Brazil, etc. is the production and commercialization of non-timber products like brazil-nuts, which are dry fruits and do not need special care for its cultivation. However, prior to its exportation it is necessary a classification process, which is done manually. This classification process is based on the number of Brazil-Nuts per pound, being grouped in large, medium, small and tiny. [1].

Actually, many researches have tried to automate the classification process, based on external damages and features such as color, size, shape or weight, all of them using computational vision and image analysis. [2], [3], [4], [5] For example, in Spain a study was presented [6], showing the use of computer vision for classification of damage in fruits, which was worked out by human inspection. They developed an approach using a computer vision system, to detect defects in the fruit peel and classify them by the type of damage. Others methodologies that allows to recognize and classify images [7], are based on color, texture and morphological features to recognize and classify horticultural products.

Also to reduce misclassification, a computer vision framework was developed in order to automatically classify the quality of corn tortillas according to five subclasses given by a sensory panel. Once the development of a feature selection algorithm is done, the most relevant features are selected for classification [8]. Likewise, the weight estimation by image analysis have been apply over different fruits such as papayas [9], citrus fruits [10], beans [11], etc.

This means, that the automation of processes involving products classification has been implemented by its high contribution to companies, because they reduce classification errors and speed up processes; but the first step in many computer vision applications is segmentation. The goal of it is to cluster pixels into salient image regions, which are called objects, and the rest of the image is known as background. It can be used for object recognition and occlusion boundary estimation without motion [12]. Then, a better segmentation process can improve the classification.

In this way, segmentation based on color spaces is a technique widely used [13], specially the YCrCb colour space has been used in image segmentation [14], [15].

In this sense, this study proposes to automate the brazil-nuts classification process according to its weight and different sizes such as Large, Medium, Small and Tiny for subsequent exportation, based on the area of each one, and a Conversion Factor calculated. YCrCb colour space was used to segment the images, with a dynamic threshold based on Colour Histograms for binarization.

The rest of the paper is organized as follows: section 2, describes the art state of classification techniques based on
computer vision, section 3 presents the proposed approach. In section 4 the automation classification process is described, section 5 shows the experiments and results and finally, section 6 draws the conclusions and future works.

II. Art State

Currently, to determine the quality of a brazil-nut, people have to make a manual process using visual inspection, which present drawbacks such as fatigue, slowness, subjectivity and many others. This is the main reason to automate the classification process. Nowadays, several kinds of classification methods have been developed such as decision tree induction [16], Bayesian networks [17], k-nearest neighbour classifier [18], genetic algorithms and fuzzy logic techniques [19], supervised and unsupervised clustering[20], [21].

In this sense, Neural Networks have been successfully applied to a variety of applications in industry. In the specific case of fruit classification, this process has been based on criterion like size, colour, shape, morphological features, texture and defects. Other methods are based on a Back-propagation Neural Network (BPNN) [2], [7]. Another case where neural networks were use was in a classification system for beans [11], which reached a performance of 90.6%. One of the main advantages of the Neural Networks is that they do not use threshold values, but need a training process, which can be a disadvantage in a real time system.

Similarly, a Fuzzy method [22] was used to classify rice grains, the inputs of this system were the area, perimeter, circularity and compactness; compared with human inspection this system reached a 90% of correct results. Likewise, a standard non-linear Bayesian discriminant analysis was used to determine the classification functions, in order to identify the type of defect in the skin of citrus fruits, [3]. This system worked with spectral information about the defects and morphological estimations; reaching a 86% of efficiency.

In the case of the quadratic analysis, most of the time gave a more accurate classification, but not significantly better than the discriminant analysis. This was showed in a color classifier for symptomatic soybean seeds [5], where the classification accuracy for linear and quadric functions ranged from 67 to 81%. Other studies [4], [8] to determine a quantitative classification algorithm for fruit shape in kiwifruit and based on support vector machine were done.

However, one of the common problems in all of these methods is the presence of noise in the image, for which new approaches have been done, [26]. One of them is the use of colour spaces or physical features, which have been studied because of their advantages, being able to segment images of fruits [27]. Finally, other techniques use features such as length, major diameter, minor diameter, mass, volume, diameter, area, eccentricity and central moments to discriminate between similar coloured defects [28].

III. Automation Classification Process

To perform the automation of the brazil-nut classification process in large, medium, small and tiny, the present paper propose to estimate the weight of each brazil-nut by using image processing, and other processes detailed below.

A. Segmentation

Considering that the weight and the area of the Brazil-Nuts have a direct relation, the first step of the approach is the segmentation process, in order to compute the area of each brazil-nut. Image 1, shows the original image, before the segmentation process. As it can be observed, the Brazil-Nuts are organized like a matrix over a light color background.

![Figure. 1: Original image](image)

The segmentation process was done using the color space YCrCb, in the Cb channel, this channel converts the image into a luminance, chroma blue, and chroma red components, instead of the RGB representation. Using the YCrCb color space is possible to determine only the intensity, or the color hue. This could allow the creation of more precise color detectors, since the color intensity is removed considering the Cr or Cb vectors. Likewise, the Cb channel was chosen because the chrominance levels in blue are higher, performing a better segmentation, respect to the Y and Cr channel. Image 2 shows the original image in the Cb channel.
After the Cb channel of the image is computed, the image is scaled in order to improve in 0.5 this channel, getting the image in gray scale. The image will then be binarized and scaled because it decreases the brightness level, establishing a greater difference between the background and the Brazil-Nuts. Image 3 shows the scaled image in the Cb channel.

The threshold used for the binarization is dynamic and is based on the color histogram. A color histogram is a representation of the distribution of colors in an image, it counts the number of occurrences of each color, allowing to identify the predominant colors in an image. Then, the threshold is computed calculating the average of the color histogram; this value will be different in each image because each image is affected by external factors such as luminosity, which cause changes in the color and tonality. Image 4 shows a color histogram and the image binarized using this approach.

As can be observed in image 4, there is still noise presence, which could be caused by brightness, shadows in the edges and others. To solve this problem, and considering that the area of the Brazil-Nuts is considerably bigger respect to the area of the noise, small areas which are outside or inside the object are deleted, as showed in image 5.

B. Features Extraction

After the image has been segmented, as is showed in image 5, the next step consists in feature extraction; in this case, the features extraction is given by the area of each brazil-nut. Image 6 shows the Brazil-Nuts correctly segmented.
covered within image, in order to calculate their area.

![Figure 7: Cb channel image](image)

Algorithm 1 describes the segmentation, feature extraction and classification process.

### C. Weight Estimation

To begin with the method, it was necessary a database with several samples of Brazil-Nuts for each Type. The proposed method to estimate the weight of the Brazil-Nuts is based on a conversion factor $F_c$, which is determined over a training set of images of the 4 types of Brazil-Nuts. The process is described as follows:

1. The first step is to obtain the real weight ($W_r$) of each brazil-nut; this was obtained using a balance of 0.01 gr. of precision. The average of each type of Brazil-Nuts ($W_{rp}$) was computed using the next equation:

\[
W_{rp} = \frac{\sum_{i=1}^{n} W_r}{n}
\]  

(1)

Where $W_{rp}$ represent the $j^{th}$ real average weight for each type of Brazil-Nuts, $W_r$ represent the $i^{th}$ real weight of each brazil-nut $i$ of each group $j$ and $n$ is the number of types of Brazil-Nuts, in this case is a constant of value 4.

2. For each brazil-nut, the area ($A_a$) and the average of each type are calculated using the equation below:

\[
A_{ap} = \frac{\sum_{i=1}^{n} (A_{a_i})}{n}
\]  

(2)

Where $A_{ap}$ represent the $j^{th}$ average area of each type of brazil-nut, $A_{a_i}$ represent the $i^{th}$ area of each brazil-nut belongs to the group $j$ and $n$ is the number of types of brazil-nuts.

---

**Algorithm 1** Segmentation, Feature Extraction and Classification Process

1: Given an image $I$;
2: Resize Image
3: //Segmentation using YCrCb Colour Spaces
4: Change the Colour Space using the Cb Channel
5: //Segmentation Image
6: for $j = 1$ to $All\_Images$ do
7: Obtaining the color histogram of each image $I_j$
8: Binarize the image with a dynamic threshold based on the color histogram
9: Binarize($I_j$);
10: Elimination of noise inside the image $I_j$
11: //Compute Areas Brazil-Nuts $A_k$ inside a image $I_j$
12: for $k = 1$ to $Total\_Areas$ do
13: Compute $A_k$ total area.
14: end for
15: end for
16: //Classification Process
17: for $j = 1$ to $All\_Images$ do
18: for $k = 1$ to $Total\_Areas$ do
19: Read $A_k$ areas of each $I_j$ image;
20: area($k$) = $Image_j.area(k)$;
21: //Conversion Factor ($F_c$)
22: Compute the weight
23: if $w(i) > 4.12$ then
24: It is Large Brazil-Nuts
25: else if $w(i) > 3.23$ then
26: It is Medium Brazil-Nuts
27: else if $w(i) > 2.83$ then
28: It is Small Brazil-Nuts
29: else
30: It is Tiny Brazil-Nuts
31: end if
32: $k ← k + 1$
33: end for
34: $j ← j + 1$
35: end for
3. The Conversion Factor \( F_c \), is a numerical factor used to make a weight estimation, it is based on the average weight \( W_{r,p} \) and the average area \( A_{a,p} \) per type of brazil-nut; in this way, it is possible to establish a correlation between the area and the weight of each group. The factor conversion is given by the next equation:

\[
F_c = \frac{\sum_{j=1}^{n} (W_{r,p_j})}{A_{a,p_j}}
\]  

(3)

Where \( W_{r,p_j} \) represent the \( j^{th} \) average weight real of each type of Brazil-Nuts,\( A_{a,p_j} \) represent the \( j^{th} \) average area of each type of brazil-nut, and \( n \) is the number of types of Brazil-nuts.

The algorithm 2 describes the calculation of the conversion factor.

**Algorithm 2 Calculation of conversion factor**

1. Read image set test
2. AverageArea=0, AverageWeight=0,
3. TotalArea=0, TotalWeight=0,
4. for \( i = 1 \rightarrow TypeBrasilNuts \) do
5. \( n=0 \)
6. for \( k = 1 \rightarrow TotalAreas \) do
7. Compute the total area,
   \( TotalArea = TotalArea + A_k \)
8. Compute the total real weight,
   \( TotalWeight = TotalWeight + wr_k \)
9. \( n=n+1 \);
10. end for
11. \( AverageArea_i = TotalArea/n; \) Equation 1.
12. \( AverageWeight_i = TotalWeight/n; \) Equation 2.
13. end for
14. AreaWeightAverage=0
15. for \( i = 1 \rightarrow TypeBrasilNuts \) do
16. //Compute the relationship between the average area and the average weight for each type of brazil-nuts
   \( A_{W_i} \)
17. \( AverageAreaWeight = AverageAreaWeight + A_{W_i} \)
18. \( n=n+1 \);
19. end for
20. \( F_c = AverageAreaWeight/n; \) Equation 3.

4. The weight estimation of each brazil-nut is based on the Conversion Factor, which give us the weight of an area in an image; in this sense, the real weight is given by the next equation:

\[
W_c = (A_{a_i}) \times (F_c)
\]

(4)

\( A_{a_i} \) represent the \( i^{th} \) areas of each brazil-nut \( i \) that can be any type, and \( F_c \) represent the Conversion Factor.

The Conversion Factor is indistinct for all types of brazil-nuts, so it does not require reformulation.

**IV. Experiments and Results**

In order to evaluate the efficiency of the proposed classification method, it has been tested on images of brazil-nuts as it is described below.

**A. Database**

The images used for the experiments were taken from the database SSCCA-5 of the Research and Development Center of the Saint Agustín National University in Perú. For these experiments a total of 1170 Brazil-Nuts: 209 large, 239 medium, 143 small and 579 tiny, were used. Each image contains 30 brazil-nuts as is shown in figure 1. These images were taken with a digital camera Canon G-9, aperture value of F8.0, shutter speed of 1/125 seconds and ISO 400 in a closed environment and controlled white illumination. These images were taken at a distance of 30 cm. as is shown in Figure 1.

For the experiments a special light box set-up was made, where the Brazil-Nuts were illuminated. The light box has one opening to allow locating the Brazil-Nuts inside the box. The opening was closed when the Brazil-nuts were inside in order to have a constant illumination. For the illumination a circular fluorescent was used. The camera was mounted on the upper part of the box, at a vertical distance of 30cm from the base.

For the selection process four Brazil-nuts types such as Large, Medium, Small and Tiny were considered. Figure 8 shows the four types of Brazil-nuts.

According to the weight \( w \), the classification by types is:

- \( 4.13 \leq w < 8.4 \) Large Brazil-Nuts
- \( 3.24 \leq w < 4.13 \) Medium Brazil-Nuts
- \( 2.84 \leq w < 3.24 \) Small Brazil-Nuts
- \( 2.06 \leq w < 2.84 \) Tiny Brazil-Nuts

**B. Segmentation and Feature Extraction**

For the segmentation process the color space used was the YCrCb in the Cb channel, reaching a 99.8% of Brazil-nuts segmented correctly, as shown in image 5. Also, like it was mentioned in previous sections, Brazil-nuts were segmented using a dynamic threshold based on colour histograms. For this reason each Brazil-nuts image has a different threshold value, even if these are the of same type. Below, in Table 1, it is present some sample data of the results.
C. Conversion Factor

The value for the conversion factor was determined using all 386 images of the training set. The result determined 0.002706 as conversion value.

D. Weight Estimation

The estimation of weight is based on two facts: the Conversion Factor and the Brazil-Nuts area. Using the equation 1 the weight of each Brazil-nut could be calculated and according to the weight obtained it is possible to define the Brazil-nut type. This is shown below in Table 2, as is also the weight and performance achieved by the results.

Table 2: Sample of Computed of Weight Brazil-nuts

<table>
<thead>
<tr>
<th>Brazil-Nuts Types</th>
<th>Real Weight</th>
<th>Estimated Weight</th>
<th>Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>5.3</td>
<td>5.0</td>
<td>94.3</td>
</tr>
<tr>
<td>Large</td>
<td>4.9</td>
<td>4.9</td>
<td>100</td>
</tr>
<tr>
<td>Large</td>
<td>4.6</td>
<td>4.4</td>
<td>95.7</td>
</tr>
<tr>
<td>Large</td>
<td>4.3</td>
<td>4.2</td>
<td>97.7</td>
</tr>
<tr>
<td>Medium</td>
<td>4.1</td>
<td>4.1</td>
<td>100</td>
</tr>
<tr>
<td>Medium</td>
<td>3.8</td>
<td>3.8</td>
<td>100</td>
</tr>
<tr>
<td>Medium</td>
<td>3.6</td>
<td>3.5</td>
<td>97.2</td>
</tr>
<tr>
<td>Medium</td>
<td>3.6</td>
<td>3.5</td>
<td>97.2</td>
</tr>
<tr>
<td>Small</td>
<td>3.2</td>
<td>3.1</td>
<td>96.9</td>
</tr>
<tr>
<td>Small</td>
<td>3.1</td>
<td>3.0</td>
<td>96.8</td>
</tr>
<tr>
<td>Small</td>
<td>3.1</td>
<td>3.1</td>
<td>100</td>
</tr>
<tr>
<td>Small</td>
<td>2.9</td>
<td>2.9</td>
<td>100</td>
</tr>
<tr>
<td>Tiny</td>
<td>2.5</td>
<td>2.5</td>
<td>100</td>
</tr>
<tr>
<td>Tiny</td>
<td>2.4</td>
<td>2.2</td>
<td>91.7</td>
</tr>
<tr>
<td>Tiny</td>
<td>2.3</td>
<td>2.3</td>
<td>100</td>
</tr>
<tr>
<td>Tiny</td>
<td>1.7</td>
<td>1.6</td>
<td>94.1</td>
</tr>
</tbody>
</table>

E. Classification Process

The results for the approach of classification process are given in Table 3. This approach of classification process achieves an accuracy of 1167 Brazil-nuts out of 1170 in the test set. For each type the accuracy was of 99.5%, 99.2%, 100% and 100% respectively. Thus, an overall performance of 99.7% was achieved on the test set. Wrong classification was observed in boundaries of types, where the difference between types is 0.4g. Estimated weights that did not match with their real type but were inside boundaries of their respective type were considered correctly classified, whereas estimated weights that were outside of boundaries of their respective type were considered misclassified. Classification and misclassification for each type are shown in Figure 3.
Table 3: Results of Classification on the test data set.

<table>
<thead>
<tr>
<th>Brazil-Nuts Types</th>
<th>Large</th>
<th>Medium</th>
<th>Small</th>
<th>Tiny</th>
<th>Total</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>208</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>209</td>
<td>99.5</td>
</tr>
<tr>
<td>Medium</td>
<td>0</td>
<td>237</td>
<td>1</td>
<td>1</td>
<td>239</td>
<td>99.2</td>
</tr>
<tr>
<td>Small</td>
<td>0</td>
<td>0</td>
<td>143</td>
<td>0</td>
<td>143</td>
<td>100</td>
</tr>
<tr>
<td>Tiny</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>579</td>
<td>579</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>208</td>
<td>238</td>
<td>144</td>
<td>580</td>
<td>1170</td>
<td><strong>99.7</strong></td>
</tr>
</tbody>
</table>

V. Conclusions

In this study it was shown that it is possible to automate the classification process of Brazil-nuts using digital images and estimating their weight, based on a conversion factor. Also, the calculated Conversion Factor was an efficient method for determining the weight of each Brazil-nut. The efficiency achieved in the Classification process for Brazil-nuts has been of 99.5%, 99.2%, 100% and 100% for Large, Medium, Small and Tiny Brazil-nuts respectively. The overall efficiency achieved by the proposal is 99.7% in images taken at a distance of 30 cm.

Likewise, the Cb channel of YCrCb colour space for the calculation of the Brazil-nut area through segmentation has allowed for the proposed method to work out properly. Certainly, to get the real Brazil-nut area in the training set yields a more accurate conversion factor, which is why this value is different for each type of Brazil-nut. The implementation of a dynamic threshold based on colour histograms in the proposed method has also been very important as there were no problems with the colour in the images caused by external influences such as lighting, brightness, etc.; obtaining a high performance in the segmentation process, 99.8% rating.

References


**Author Biographies**

**Sonia Castelo Quispe** was born at Perú in 1988, received the Bachelor’s degree in System Engineering from Saint Agustin National University in 2010. She is a researcher in several areas including Image Processing and database managements applied to the Industry in Saint Agustin National University Research and Software Development Center, Cathedra CONCyTEC in TIC’s. She is interested in image processing, computer vision, database managements, and artificial intelligent.

**July Diana Banda Tapia** was born in Arequipa, Perú in 1989. She finished her studies in Computer Science on December 2011 at the Saint Pablo Catholic University in Arequipa-Peru and received the bachelor’s degree. She works in the Research and Software Development Center of the Saint Agustin National University - Cathedra CONCyTEC in TIC’s. Her main interests are bioinformatic and image processing.

**Mónica Nora López Paredes** was born in Arequipa, Perú in 1982. She finished her studies in Computer Science on December 2009 at the Saint Pablo Catholic University in Arequipa and received the bachelor’s degree. She works in the Research and Software Development Center of the Saint Agustin National University - Cathedra CONCyTEC in TIC’s. Her main interests are artificial intelligent and image processing.

**Dennis Barrios Aranibar** He is professor at Saint Pablo Catholic University and researcher at Software Development Center of the Saint Agustin National University - Cathedra CONCyTEC in TIC’s. He received is doctoral degree at Universidad Federal do Rio Grande do Norte (2009) and master degree at the same university (2005). His main interests are robotic and automation applied to industry, multiagent systems, and machine learning. He is President of Computing Peruvian Society and Peruvian Representative to the Latin American Studies in computer Science.

**Raquel Patiño Escarcina** Graduated in Systems Engineering at the Saint Agustin National University in Peru, Master in Systems and Computation(2005) and Doctor in Electric Engineering(2009) at the Federal do Rio Grande do Norte University, in Brasil. She works as professor and researcher at the San Pablo Catholic University and in the Research and Software Development Center of the Saint Agustin National University - Cathedra CONCyTEC in TIC’s. Her main interests are computer vision, image processing and artificial intelligence.