**Chapter 3**

**METHODOLOGY**

**Introduction**

The Philippine Grain Standardization Program is a government program spearheaded by the National Food Authority to provide commercial assessment standards for the determination of the grade and quality of milled rice products. The implementation of the program started on September 21, 2002. From its establishment, the National Grains Standard has been formed. The National Grains Standards defined the characteristics classification of the rice grain samples. Factors for determining grade include dimensional length, degree of milling, percentage by weight of broken kernels, brewers, red kernels, immature kernels, chalky kernels, damaged kernels, yellow kernels, age-related changes, and other characteristics.

The grades are based on the percentage by weight of the classified grains to the overall weight of the product. The Grade 5 is the lowest and the Premium grade is the highest grade a milled rice product can be classified to. Moreover, the implementation of these standards in the market is expected to boost the quality of the rice products in the Philippines.

The purposes of this study are to lessen the subjectivity of rice grading assessments based on the National Grain Standards by utilizing the consistency and precision of computer-aided assessments and to speed up the grading time. Using image processing methodologies, the study aims to develop a milled rice grading system that is portable and accessible to people and organizations who are working directly on rice like millers, distributors, and farmers. The main parts of this study are directed towards the gathering of the qualitative reference values set by the NGS and the rice quality assessors; creation of a portable standalone device equipped with image analysis software that grades the rice samples based on an image of non-overlapping grains; and the display of the grade report. Activities performed by the image processing application include the counting and dimensioning, color analysis, grain classification, averaging, and grading.

Figure. The Conceptual Framework

The portable automated grading system grades the rice grain samples based on their features and characteristics. The input to the system will be an image of non-overlapping rice grains obtained from a special image acquisition platform designed to minimize variation in lighting effects and illuminated by a constant setting light source. The Otsu’s algorithm will be applied to find the optimum threshold value to separate the grain pixels from the background pixel. Pixels of the grain will retain their original values while background pixels will be converted to black pixel values. Furthermore, using connected component labelling, the representative pixels will be labelled into neighbors with each neighbor representing a single grain. The labels act as counting mechanism. For each grain, a label will be assigned (e.g. 1, 2, 3, 4, etc.). The grain count will be obtained during this process. The individual analysis of grains will begin afterwards.

For every single grain, morphological dimensioning will be performed to measure the major axis length. After dimensioning, the color analysis will be performed. Before initiating the color analysis, the color space will be converted to the CIE La\*b\* color space to facilitate better analysis due to the reason that CIE La\*b\* is deemed to be perceptually uniform \*\*\*\*ref here\*\*\*\*. Each grain in the sample will be classified. The grain classification process has two parts: (a) grain size classification and (b) grain type classification. In the grain size classification, the size of the grain is compared to the maximum length of the grains of the sample. The system will classify the grain into broken, brewer, or unbroken. In the grain type classification, the system will classify the grain into immature, red, yellow kernel, or two or more of the mentioned classifications based on their comparison to the reference values. The system will maintain a count for every classification.

After classification, chalkiness and bran presence will be assessed for each grain. The areas of the bran and the chalky regions are the region of interest (ROI) in each grain. Each area will be compared to the overall area of the grain. The grain will then be classified as chalky or not and the percentage area of the bran presence will be recorded as it indicates the degree of milling.

The number of grains classified to grain type classifications (i.e. red, immature, yellow, etc.) will be counted. The average length of the grain sample is the average length of the grains classified as unbroken. The percentage by weight of the immature, red, yellow, and chalky kernels are computed based on the ratio of each of the count to the total number of grains in the sample. The degree of milling is obtained by regional area percentage of the bran.

The data obtained from the image processing methodologies are compared and classified into grades based on the National Grain Standards. The system will display these determining factors along with the grade of the rice grain samples. The grade of the samples ranges from Grade 5 being the lowest to Premium as the highest grade a sample can be classified to.

**Quantification of The Qualitative Characteristics to Be Used for Reference**

With the traditional method of grading rice, the qualitative factors (e.g. immaturity, yellowness, etc.) rely on the expertise and perception of the assessor. Moreover, the National Grain Standards provided qualitative measures for assessing a rice grain. An example of this is the chalkiness of a grain. The NGS defined the color of the chalky region as ‘white as a white chalk’. Before doing the image processing, the system needs to obtain the reference values that would indicate the mentioned qualitative characteristics.

The following factors are qualitative in nature and the references are based from the information obtained from the NGS and a certified grade assessor: (a) degree of milling, (b) immaturity, (c) chalkiness, (d) redness, and (e) yellowness. Since the CIE La\*b\* color space is designed to be similar to how human vision perceives colors and light, and the image acquisition platform is set to have constant lighting settings, reference colors can be recorded from NGS compliant reference images.

Table. summarizes the qualitative characteristics that need to be quantified along with their description. The descriptions are provided by the NGS. For each classification type, a 300-grain sample size will be used to get the reference values.

Table. The qualitative characteristics that must be quantified

|  |  |
| --- | --- |
| Classification | NGS Description |
| Immature Kernel | The immature kernels are those, whole or broken, which are light green and chalky with soft texture. |
| Chalky Kernel | The chalky kernels are those, whole or broken, one-half or more of which is white like the color of white chalk and is brittle upon removal of the hull for palay. |
| Red Kernel | The red kernels are those that have red bran covering, wholly or partly. |
| Yellow Kernel | The discolored kernels are kernels that have changed their original color as a result of heating and other means. |

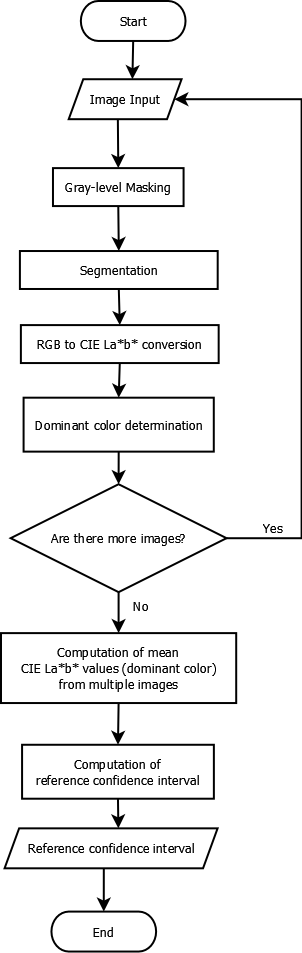


Figure. The process flowchart of the determination of reference confidence interval

Figure is the process flowchart for the determination of the reference confidence interval. Instead of having an individual CIE La\*b\* color as a reference value, an interval was used to statistically represent the population. Multiple reference images are input to the system in training mode. A segmentation that uses Otsu’s method to find the optimum threshold value will be performed to separate the background from the foreground (grains). Before performing Otsu’s method, a gray-level mask is obtained by duplicating the original image, *O*, and converting it into a gray-level image . The Otsu’s method will be applied to the and it will be binarized into . Based on the threshold, the background pixels of will be set to have a value of (0,0,0) and the foreground pixels will have (1,1,1). {\displaystyle A\circ B} The is a mask to be applied to the image *O*. An element-wise multiplication will be applied to image *O* and resulting to a threshold image In this way, all the background pixels of will have a value of black and the foreground pixels will retain their color values. The process is described by the equation \_\_\_\_\_\_.

After segmentation, the color space will be converted into the CIE La\*b\* color space. The segmented image *,* is defined in equation \_\_\_\_\_\_\_\_\_\_\_\_\_\_:

Where R = red channel value

G = green channel value

B = blue channel value

m = width coordinate

h = height coordinate

Whereas, the converted image *I* is defined as:

Where l = lightness value

a = green—red value

b = blue—yellow value

m = width coordinate

h = height coordinate

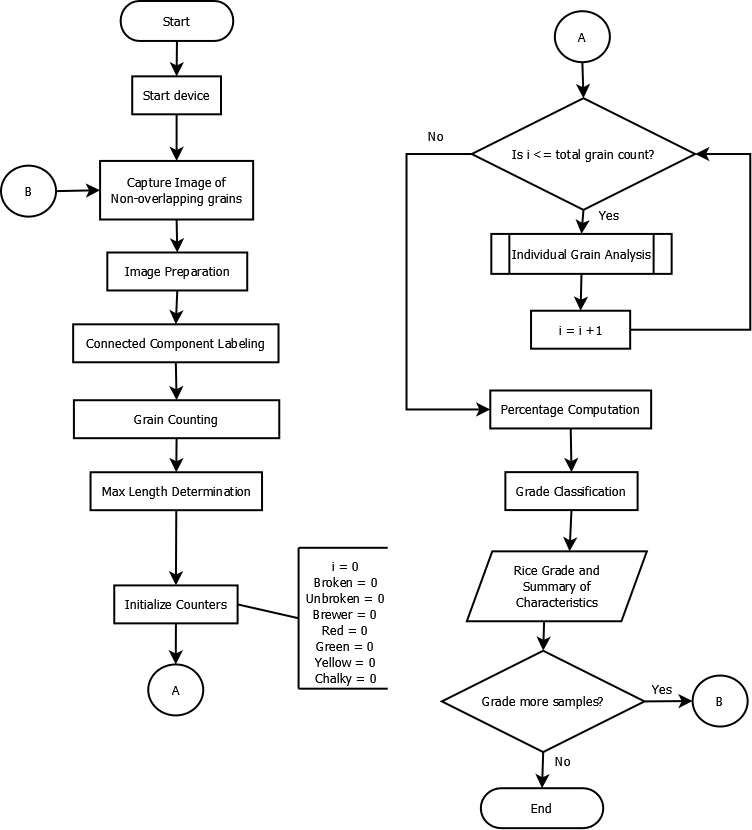
After the color space conversion, the dominant color will be obtained. To obtain the dominant color of the reference image, color histograms are created. For every image, there will be a color histogram for each CIE La\*b\* color space dimension (a total of 3 for every image: l, a, and b histogram). Each histogram will have 201 (-100 to 100) bins that contain the number of pixels for every dimension value. The bin with the highest number of pixel represents dominant color. Note that the background pixels will be ignored. Therefore, they don’t have effect to the determination of the dominant color. For every color histograms (i.e. l, a, and b), the dominant bin is considered. A CIE La\*b\* color value is comprised of the L, a\*, and b\* bins with the most number of pixels, . This process will be done on every reference images. A total of *n* reference values from *n* reference images comprising of 300 grain sample for each classification type will be the result of the entire process. Until all reference images are processed, the dominant color determination will stop. All the obtained reference values with each described as a point (l, a, b) will be averaged. The value is represented by *E* in the equation.

An aggregate confidence interval, *C*, with a confidence level of α = 95%, sample size *n*, means *l, a,* and *b*, standard deviation , confidence coefficient (using z-test) can be described by the equation.

Where

The *C* interval will be used as a model for the qualitative characteristics of red, immature, white region of the chalky, and yellow kernels.

**The Grading Process**

****

**Figure.** The flowchart for the grading process

**Image preparation**

Figure illustrates the grading process using the device. An image of non-overlapping grains more than \_\_\_\_\_\_\_\_\_ is captured after the device is turned on. Unlike the process of getting the reference images, the grains must have the same upright position as illustrated in figure \_\_\_ to facilitate more effective dimensioning process. The image preparation begins with the segmentation process. The Otsu’s method will also be used to determine the optimum threshold value that will separate the background from the foreground. The equation \_\_\_\_ from the process of segmentation of reference image will be used to segment the image. The preparation process produces an image with pixel values in CIE La\*b\* color space. The segmented image can be represented by as with the process described by the equation \_\_\_\_\_\_\_.

**Connected-component labelling, counting, and maximum length determination**

The segmented images will undergo connected-component labelling. In this process, the aggregate of colored pixels will be clustered together and a bounding box be formed around them. These groups of pixels represent the individual grains. Another mask image, , is constructed from the original image, , following the connected component labelling. The image contains the labelled blobs of the grains. It serves as a duplicate of but it has been transformed to contain the coordinate and label of the connected components. The connected components will be labelled with numbers 1 to *n* with *n* being then number of the grains in the sample. The grain count in the sample is recorded to be used later. After labelling, a minimum bounding-box will be constructed and be wrapped around the individual grains. The area of the minimum bounding-box will be the approximate area of the grain. However, the interest of this study is to measure the major axis length as the NGS defined all the size considerations and classifications based on it. All the major axis lengths of the grains will be measured and the biggest length will be recorded.

**Grain Type Classification**

The gain type classification is the process of counting the number of grains classified as (1) immature/green, (2) yellow, (3) red and (4) chalky kernel. Each of the labelled grain components will be subjected to this classification process. For the first three classifications, the dominant color for each grain will be the basis. The individual grain analysis determines the dominant color of the sub-image like the process done in determining the reference image. Once the dominant color for the grain has been determined, the color difference will be obtained. The 1976 CIE La\*b\* color space’s color difference metric is called the as defined by the equation \_\_\_\_\_\_\_\_. The color difference is a Euclidean distance between two CIE La\*b\* color values.