INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Article | DOI: 10.21307/ijssis-2019-010

Issue 1 | Vol. 12 (2019)

Image processing approach for grading tobacco leaf based on color and quality

Agus Harjoko[1], Adhi Prahara[2], Tri Wahyu Supardi[1], Ika Candradewi[1], Reza Pulungan[1] and Sri Hartati[1]

[1]Department of Computer Science and Electronics, Universitas Gadjah Mada, Sekip Utara, Bulaksumur BLS 21, Yogyakarta, Indonesia.

[2]Departement of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia.

E-mails: aharjoko@ugm.ac.id, adhi.prahara@tif.uad.ac.id, twsupardi@ugm.ac.id, ika.candradewi@ugm.ac.id, pulungan@ugm.ac.id, shartati@ugm.ac.id

This paper was edited by Subhas Chandra Mukhopadhyay.

Received for publication January 30, 2019.

Abstract

Implementation of automation technology for grading tobacco leaf was very promising. In Indonesia, grading tobacco leaf was done manually and relied on the skill and experience of tobacco leaf graders. Large tobacco plantation needed many graders, and the workers needed to be trained, to become a skilled grader. It would take a long time and substantial cost to prepare sufficient graders. Even if the plantation had enough graders, monotonous and long duration of work would raise the human error. Therefore, we proposed a method for grading tobacco leaf based on color and quality using image processing techniques. This work covered quality inspection of tobacco leaf, namely leaf defect detection and classification of tobacco leaf based on color. Image processing techniques such as image thresholding, morphological operation, blob detection, and color analysis of tobacco leaf were employed to determine the grade of tobacco leaf. From the experiment, the proposed method was able to detect a leaf defect and able to classify tobacco leaf with 91.667% accuracy.

Keywords
Tobacco leaf, Grading system, Image processing, Color, Quality.

Tobacco (*Nicotiana tabacum*) is believed to have originated from South America (FAO, 2018a). The cultivation of tobacco leaf then spread globally within decades after Columbus helped to import tobacco into Europe in the 15th century. This event marks the starting point of tobacco leaf cultivation for commercial purposes. The mechanization of cigarette manufacturers in the 1880s also helps the growth of the cigarette market thus increasing demand for tobacco leaf (Drope and Schluger, 2018). Tobacco cultivation now is one of the most profitable agricultural commodities in the world.

In Indonesia, tobacco cultivation occupies 0.37% of total agricultural land. The production reaches 196,154 tons of tobacco in 2016. That what makes Indonesia ranked fifth in the world (FAO, 2018b). However, the gap in the production of tobacco is too far compared to China with 2,805,615 tons or India with 761,318 tons. One of the reasons is the tobacco leaf grading system in Indonesia still done manually, relies on the skill and experience of tobacco leaf graders. This case is found in PT. Perkebunan Nusantara X (PTPN X, 2016), one of the tobacco manufacturer company in Indonesia. PTPN X plants Virginia type of tobacco leaf to target the cigarettes export market. The grading system classifies tobacco leaf into three categories with eight classes per category and done manually by the workers. The decision made by the workers is subjective, depend on their skill and experience, which often leads to false classification. Moreover, the lighting condition on its grading chamber relies on sunlight that susceptible to illumination changes.

Large tobacco plantation needs many tobacco leaf graders. In order to become a skilled grader, the worker needs to be trained. The training will take a long time and substantial cost to produce sufficient graders. Even if the plantation has enough graders, the monotonous and long duration of work will raise the human error.

The high demand for tobacco leaf triggers the mechanization of tobacco manufacture to push mass production. Therefore, many types of research that support tobacco manufacture especially the tobacco leaf grading system are conducted and sponsored by agribusiness companies. Most of the research works employ image processing techniques to determine the quality of tobacco leaf (Ni et al., 2009; Zhang and Zhang, 2011; Liu et al., 2012; Mallikarjuna and Guru, 2013; Bin et al., 2016; Jianqiang et al., 2018). It was expected because image processing techniques have been widely used in agriculture to processing post-harvest agricultural products (Chen et al., 2002; Laykin et al., 2002; Mashithoh, 2013).

In the study of Zhang and Zhang (2011), image processing techniques were used to extract and analyze color, size, shape, and surface texture on a flue-cured tobacco leaf. The tobacco leaf grading system used fuzzy comprehensive evaluation and resulted in 72% classification accuracy. In the study of Mallikarjuna and Guru (2013), a feature-level fusion was proposed for the classification of tobacco leaf. They used Haar wavelets, and gray-level local texture patterns (GLTP) that fused using concatenation rule then applied sequential backward selection (SBS) to select the discriminative texture features. K-nearest neighbor was used to classifying the tobacco leaves into viz, ripe, unripe, and over-ripe.

Liu et al. (2012) used image processing and generalized regression neural networks for tobacco leaf grading. They utilized digital image technology including near-infrared spectrum to extract area, perimeter, length, width, colors as a set of features that determined tobacco leaf quality. The features were classified using a generalized regression neural network into several grades. Near-infrared (NIR) spectroscopy for tobacco leaf classification also used in the studies of Ni et al. (2009), Bin et al. (2016), Jianqiang et al. (2018). NIR spectroscopy was rapid, non-destructive, and straightforward for tobacco leaf. NIR spectra were used to predict the content of some chemical components which are essential indicators to determine the tobacco leaf quality.

This research employs image processing techniques such as image thresholding, morphological operation, blob detection, and color analysis for grading tobacco leaf. This work focuses on grading tobacco leaf based on color and quality. The method covers the quality inspection of tobacco leaf, namely, leaf defect detection and classification of tobacco leaf based on color. The rest of this paper is organized as follow: Section “Tobacco leaf grading system” presents the proposed tobacco leaf grading method, Section “Image processing approach for grading tobacco leaf” presents the results and discussion, and Section “Result and discussion” presents the conclusion of this work.

Tobacco leaf grading system

Indonesia has a standardized grading system for tobacco leaves. Tobacco manufacture company such as PTPN X trains their graders according to the manual (PTPN X). The standard procedure of grading tobacco leaf consists of:

1. grading based on leaf position;
2. grading based on color and quality;
3. grading based on usage; and
4. grading based on a detailed color level.

Figure 1 shows the illustration of tobacco cultivation. The lower tobacco leaf can be harvested when the flowering started, and the upper leaf can be harvested afterward.

The leaf position is used to grade tobacco leaf as shown in Figure 2, namely (A) KOS, (B) KAK, and (C) TNG. KOS category consists of a lower tobacco leaf that grows near the ground field. The leaf is broad with the rounded tip, thin, and has smooth leaf veins. KAK category consists of middle tobacco leaf. The leaf is full of a slightly narrowed tip, slightly thick, and has smooth leaf veins. TNG category consists of upper tobacco leaf. The leaf is wide with narrowed tip, thick, and has slightly broken leaf veins.

After tobacco leaf is grouped based on leaf position, abnormality detection is applied to find diseases, stripes, and spots on the leaf surface and defect detection are applied to find holes or torn trace on the leaf surface. The leaf defects are break down into four categories based on the holes or torn position as shown in Figure 3, namely (A) top, (B) bottom, (C) side, and (D) center. This leaf defects category is one of the criteria to grade tobacco leaf based on color and quality.

The next step is to classify tobacco leaf based on maturity level and color. The maturity level of tobacco leaf is categorized as MT for young/unripe leaf, T for mature/ripe leaf, and TT for the old/over-ripe leaf. In this research, we focus on the mature/ripe category. The mature/ripe category is broken down into three sub-categories based on leaf color as shown in Figure 4, namely (A) K (yellow), (B) M (brown), and (C) B (green). This category is one of the criteria to grade tobacco leaf based on color and quality.

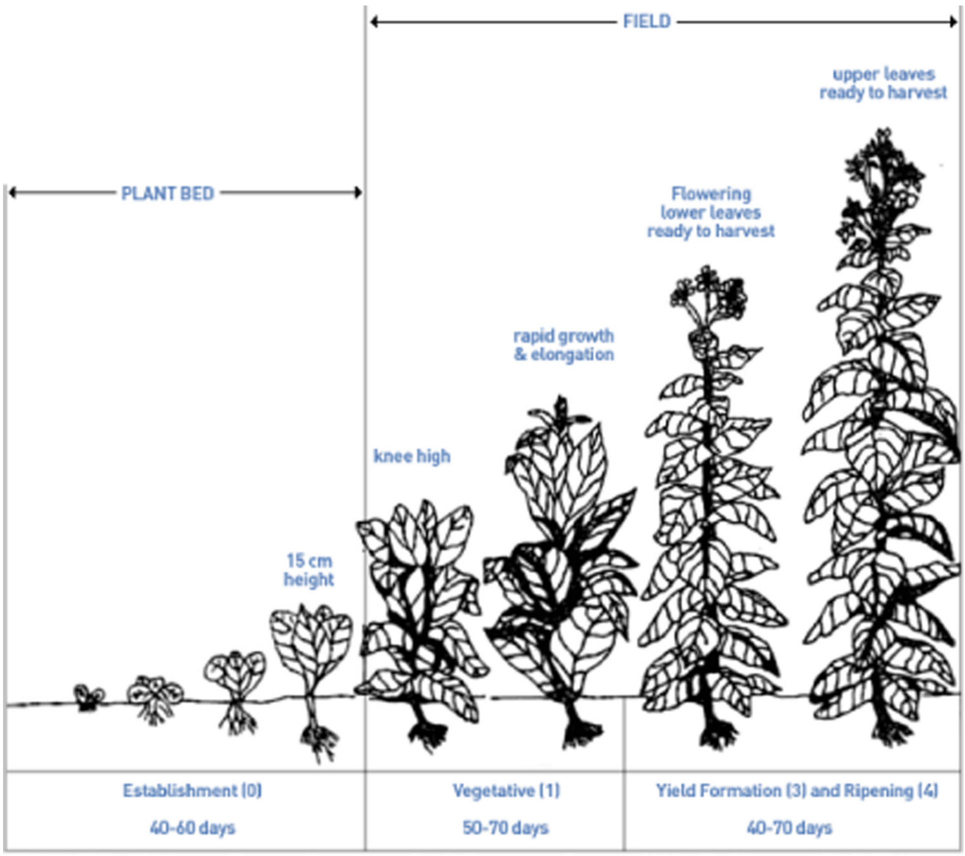

Figure 1. Illustration of tobacco cultivation (FAO, 2018a).

Image processing approach for grading tobacco leaf

In this research, image processing techniques are applied for grading tobacco leaf based on color and quality. The proposed method consists of tobacco leaf image acquisition, leaf area segmentation, leaf defects detection, and color-based tobacco leaf classification. The procedure is explained in the following steps:

1. Tobacco leaf image acquisition
   Tobacco leaf is captured using a camera in a container with a fixed resolution, standardized lighting,

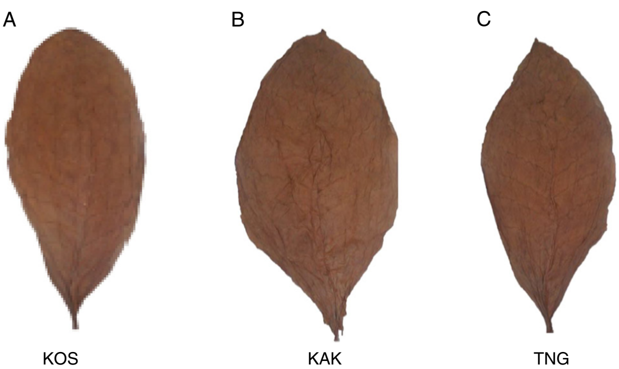

Figure 2. Tobacco leaf categories based on leaf position.

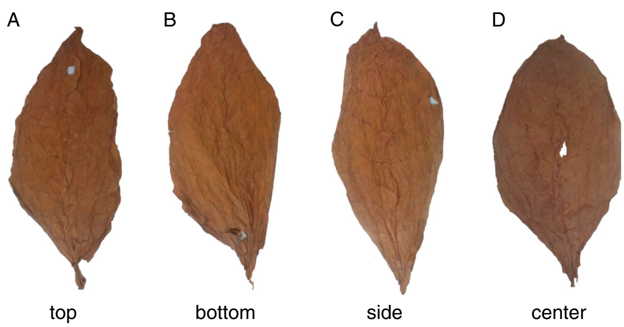

Figure 3. Leaf defect categories based on holes or torn positions.

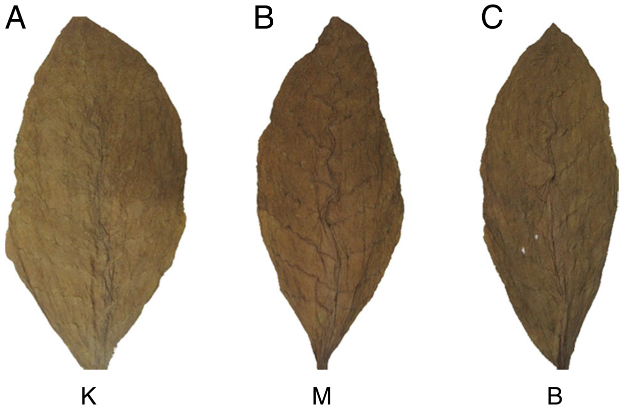

Figure 4. Mature/ripe leaf subcategories based on the color of tobacco leaf.

and uniform colored leaf bed. The camera is configured manually to turn off automatic settings such as autofocus, autoexposure, and auto white balance. The lighting is positioned vertically above the leaf bed and configured to produce constant light and to spread the light evenly. The leaf bed color is dark gray and made using material that minimizes the light's reflectance.

2. Leaf area segmentation
   Leaf area segmentation uses image thresholding in hue channel, blob detection, and morphological operation. The procedure to segment the leaf area is shown in Figure 5 and explained in the following steps:
   - The tobacco leaf image is converted to HSV color space.
   - The tobacco leaf area is segmented using image thresholding in the hue channel with a given threshold. The threshold is obtained from the experiment.
   - Blob detection is performed on the binary image of the leaf area to extract blob's contour, position, and orientation. The blob detection algorithm is set to retrieve only external contours and reject the inner contours as we were only interested in the largest blob area.
   - The blob with the largest area is designated as potential leaf area.
   - Leaf shape is rotated according to the tilt degree of leaf major axis to fix the leaf orientation.
   - The morphological operation is applied to fill the holes or torn on the leaf surface.
   - The leaf image is cropped to fit the leaf area and remove the background.
   - Leaf geometry such as size and dimension is calculated.
   - Finally, the result of leaf area segmentation is a leaf area image with normalized leaf orientation and leaf geometry information.
3. Leaf defects detection
   The defect on the leaf surface reduces the quality of tobacco leaf. Leaf defects are categorized based on the defect location on the leaf surface, namely top, side, center, and bottom category as shown in Figure 6. From Figure 6, the center area has ellipse shape which proportional to the leaf size, the top part is on the leaf tip, the bottom part is on the leaf base, and the rest belong to side category.
   The procedure to detect leaf defects using image processing techniques is explained as follow:
   - The morphological operation is applied to fill the leaf area.
   - The difference between the filled leaf area and the original leaf area is computed to find holes. The holes are the potential of leaf defects.
   - Blob detection is performed to find contour, area, center, and position of the defects. We simplify the blob detection algorithm to retrieve

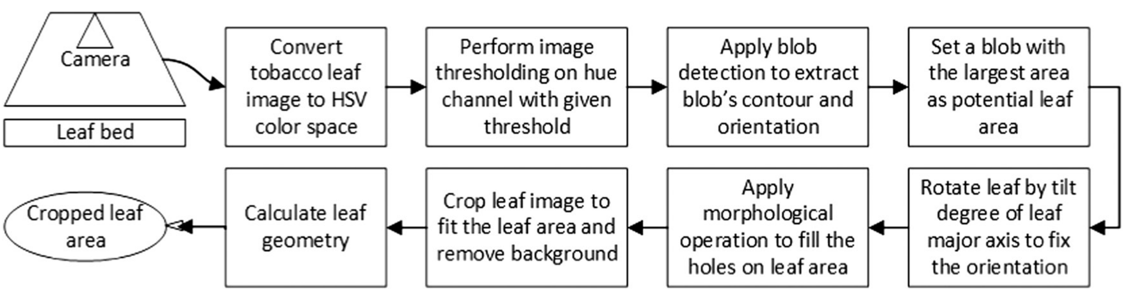

Figure 5. The procedure of leaf area segmentation.

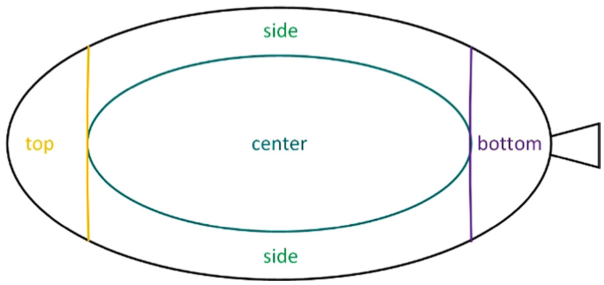

Figure 6. Area of leaf defects category.

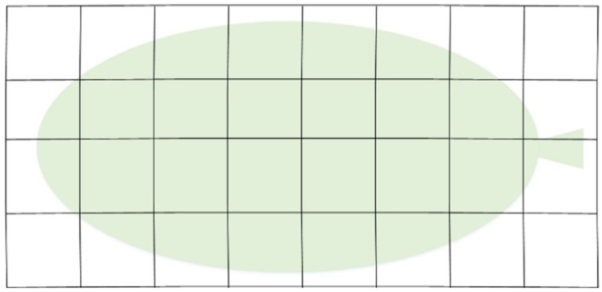

Figure 7. Small patches applied on the leaf area.

only external contours and reject the inner contours.

- Ignore all blobs that have an area smaller than a given threshold since it will not affect the quality of the tobacco leaf. In this research we only measure the quality of tobacco leaf from the size and position of leaf defects that represent by the appearance of holes on the leaf surface. The holes are detected from the difference between the leaf bed and leaf surface. Therefore, it will reduce the false detection of holes that detected as spot or disease.
- For every defect found:
  - Check if the defect center point is inside the area of the center category using (1) where (x, y) is the defect center point, (h, k) is the ellipse center point, r_x is the semi-major axis, and r_y is the semi-minor axis. If true, the defect belongs to the center category:

$$\frac{(x-h)^2}{r_x^2}+\frac{(y-k)^2}{r_y^2}\leq 1. \tag{1}$$

  - Otherwise, check if the defect center point is in the top area or the bottom area. If true, the defect belongs to the top or bottom category.
  - Otherwise, the defect belongs to the side category.
- The size of the defect is calculated to know the level of damage on the leaf surface.

4. Color-based tobacco leaves classification

In the mature/ripe category, tobacco leaf is classified into three categories based on the color of the tobacco leaf. The category is determined using a set of thresholds in the hue and value channel. The procedure to classify tobacco leaf into K, M and B category is explained as follows:

- The tobacco leaf image is converted to HSV color space.
- The image is divided into small patches with fixed size as illustrated in Figure 7. We divide the tobacco leaf into small patches to handle color diversity that will likely affect classification using a global threshold.
- The weight of each patch is calculated by dividing the leaf area inside the patch with the total number of pixels in the patch.
- The average pixel of hue and value channels in each patch is calculated.
- A set of the threshold is applied to classified each patch into a local K, M, and B category based on average hue and value. The illustration is shown in Figure 8 where Th_{low}, Th_{mid}, Th_{up} are the lower, middle, and upper threshold of hue pixel and Tv_{low}, Tv_{mid}, Tv_{up} are the lower, middle, and upper threshold of value pixel, respectively.
- The leaf color category is determined from the highest vote of color categories. The vote is calculated from the sum of the weight of each patch.

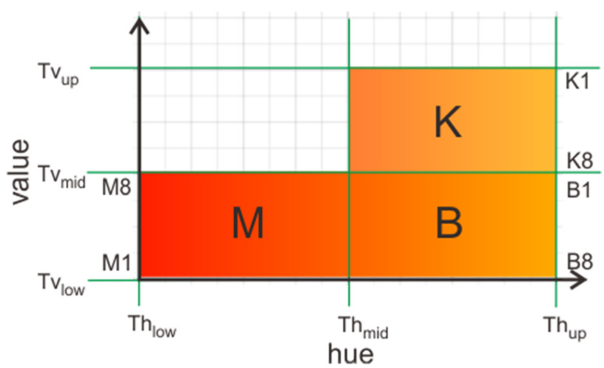

Figure 8. A set of the threshold to determine the color category.

Figure 9. A set of the threshold to determine the class in each color category.

- Tobacco leaf on each color category is divided into eight classes from 1 to 8 (bright to dark). The classes are classified using a set of threshold, namely Tv_0, Tv_1, …, Tv_8 from the average value channel on the leaf area as illustrated in Figure 9.

Result and discussion

The proposed tobacco leaf grading method is simulated in Matlab and runs on Intel Core i-7 6700K processor, 16 GB of RAM, and NVidia GTX 1050. The result, discussion, and parameters used in the experiment are explained as follows.

In order to grade tobacco leaf based on color and quality, tobacco leaf needs to be segmented from the background. Image thresholding, blob detection, and morphological operation are employed to segment the leaf area. From observation, the range of hue pixel in the tobacco leaf image is $15 \leq$ hue pixel ≤ 80 in the range of 0 to 180 degrees (normally is 0-360 degrees but it is converted to fit 1-byte size for each pixel). Figure 10 shows the hue value that used as a threshold to segment leaf area.

The result of leaf area segmentation is shown in Figure 11 where (A) is the input tobacco leaf image, (B) is the result of image thresholding on hue channel, (C) is the result of normalized leaf orientation, and (D) is the result of final leaf area. From Figure 11B, the ellipse approximation of the tobacco leaf area shows a tilted orientation. By rotating the leaf according to the degree of tilted on leaf major axis, the orientation is normalized (see Figure 11C). The leaf orientation needs to be normalized to provide an easier

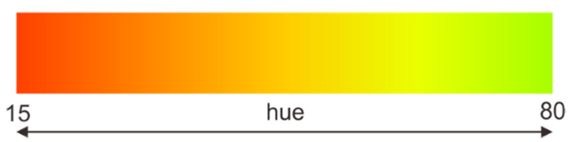

Figure 10. The threshold of hue pixel to segment leaf area.

estimation of defect position and leaf division into 10×6 small patches.

After tobacco leaf area segmentation, leaf defects detection is performed to determine the defect location on the leaf surface. The method uses morphological operation, blob detection, and blob analysis. Based on the leaf defects category, two parameters, namely, defect location and level of damage are used to determine the tobacco leaf quality. The level of damage is calculated from the size of the defect area divided by the whole leaf area. The larger level of damage means less quality of tobacco leaf. Figure 12 shows the result of leaf defects detection where (A) is the tobacco leaf image, (B) is the binary image of the original leaf area, (C) is the binary image of filled leaf area, and (D) is the result of leaf defects detection. From Figure 12D, a hole is located inside the area of center category (cyan ellipse marker) with 0.29432% level of damage. The method to detect defect on tobacco leaf is tested on several samples of defected tobacco leaf and shows 0.0 of the log average miss rate and 1.0 of average precision rate which means our proposed method identifies the defect on tobacco leaf correctly.

In order to classify tobacco leaf based on color, the tobacco leaf image is divided into small patches. Small patches are used to handle the non-uniform color of tobacco leaf produces by abnormal leaf surfaces such as diseases, stripes, or spots. Either large or small patches have their own contribution or vote (weight) to the classification. Since weight is the ratio between leaf surface area that covered by the patch and the size of the patch, the choice of patch size will not significantly affect the final classification. We perform some experiments with variation of patch sizes such as 8×4, 10×6, and 12×8. The result shows that 10×6 patch size achieved the highest classification accuracy with small margin compare to the other patch size. Small patch size means larger area inside the patch while large patch size means a smaller area inside the patch. It will affect the strength of local features used for classification.

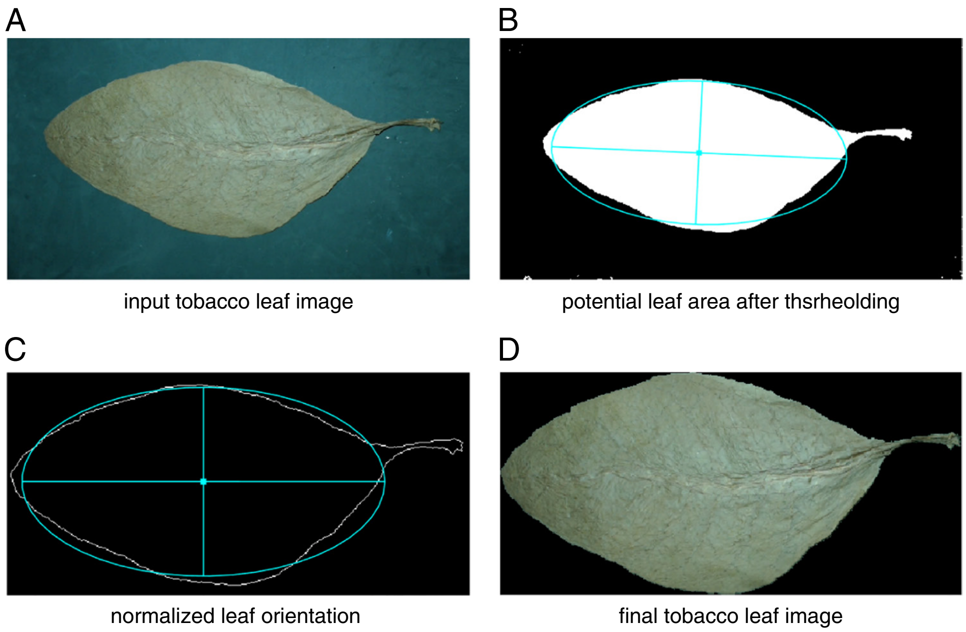

Figure 11. The result of a tobacco leaf area segmentation.

The average pixel of the hue and value channel in each patch is calculated to be classified into local K, M, B category using a set of threshold (see Figure 8). Table 1 shows the threshold derived from analysis of hue-value features distribution of SL (Special Lot)-type tobacco leaf training data. Figure 13 shows the scatter plot of the distribution of the hue-value features where red dots are

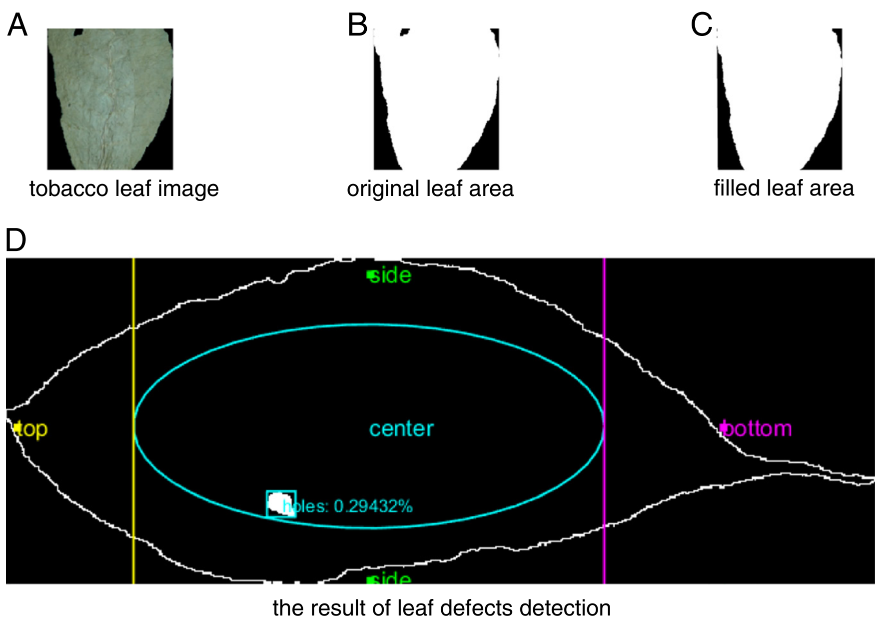

Figure 12. The result of leaf defects detection.

Table 1. The threshold for local color-based classification.

Threshold	Hue	Value
Lower	15	60
Middle	34	127
Upper	80	200

the distribution of M category, yellow dots are the distribution of K category, and green dots are the distribution of B category. The features distribution of each category is mostly clustered, make it easier to classify using simple threshold. We also consider using experimental threshold rather than automatic threshold because from the observation in tobacco manufacture company, a different type of tobacco leaf requires a different set of the threshold. According to PTPN X, the color of tobacco leaves that belong to the same type can be vary, depends on the harvest time or weather condition. Moreover, the set of threshold can be obtained using a small number of training data from the boundary of each category and more comfortable to configure and standardized by the operator.

The local color category is calculated on each small patches using the set of thresholds in Table 1. We use the voting system to determine the global color category from the local color category of each patch. Each patch contributes to the voting system based on its weight. Figure 14 shows the result of tobacco leaf image classification by majority voting (A) tobacco leaf is classified to M category with 34.3444 votes of M category from total 37.1511, (B) tobacco leaf is classified to K category with 28.775 votes of K category from total 37.4557, and (C) tobacco leaf is classified to B category with 28.9666 votes of B category from total 35.8633.

The proposed method tested on 204 tobacco leaves of SL (Special Lot) type. The classification result is shown in Table 2 using a confusion matrix. The accuracy, precision, sensitivity, and specificity score are calculated to measure the performance. The proposed method achieves 91.667% accuracy, 91.667% precision, 89.519% sensitivity, and 95.751% specificity score. False-positive occurs on the border of every category or near the threshold, especially in K and B category. From Figure 8, the only difference between K and B category is the value pixel, and it is sensitive to illumination. However, due to the standardized lighting environment that used in the tobacco leaf image acquisition process and by using a set of threshold that shown in Figure 8, the proposed method can easily handle color gradation to grade tobacco leaf based on detailed color level.

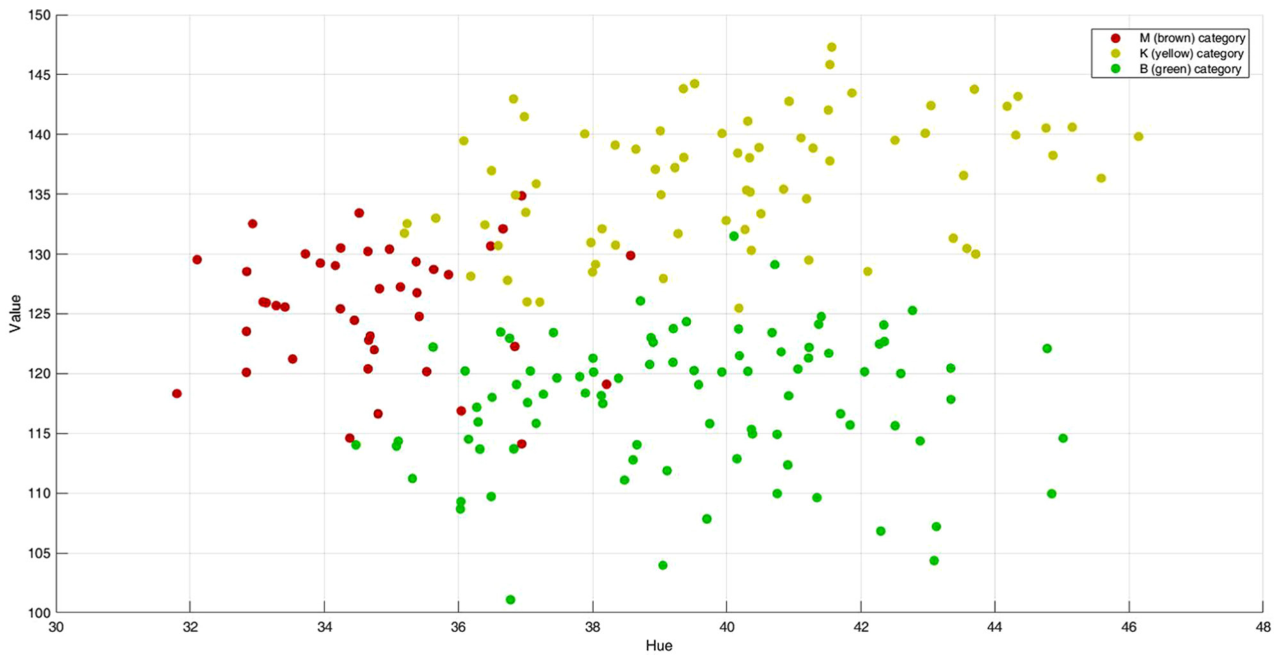

Figure 13. The distribution of hue-value features of a color-based tobacco leaf category.

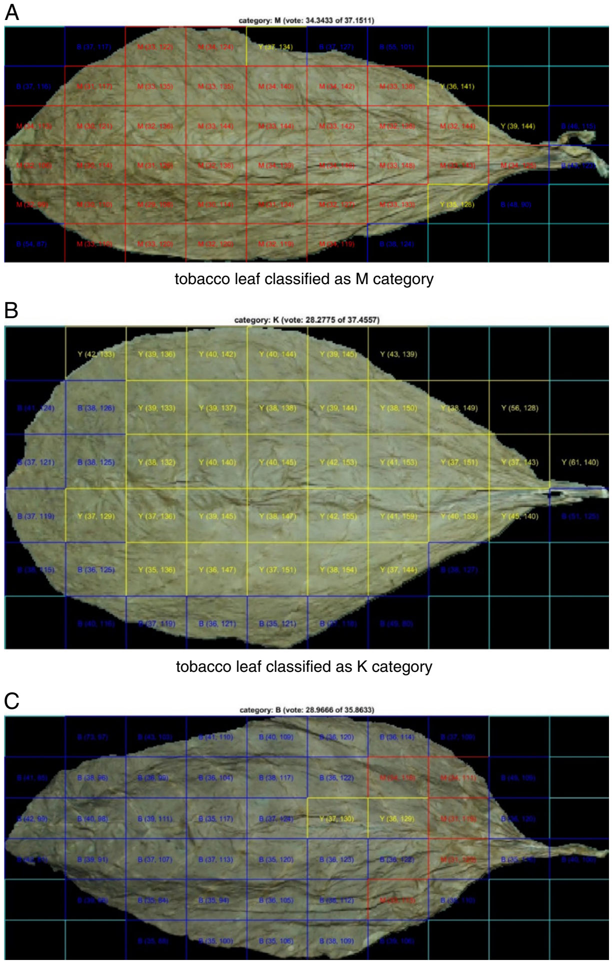

Figure 14. The result of color-based tobacco leaf classification.

Table 2. Performance evaluation of color-based tobacco leaf classification.

		Actual		
		M	**K**	**B**
	M	33	1	2
Predicted	K	8	72	4
	B	2	0	82
Accuracy	0.91667			
Precision	0.91667			
Sensitivity	0.89519			
Specificity	0.95751			

Conclusion

This research proposes a tobacco leaf grading method based on color and quality. The method employs image processing techniques such as image thresholding, morphological operation, blob detection, blob analysis, and color analysis. The result shows that the proposed system can accurately classify tobacco leaf with 91.667% accuracy into K (yellow), M (brown), and B (green) color categories. The classification method uses a set of the threshold to determine the local color category in small patches of the tobacco leaf image. The global color category is determined using majority voting from each patch. For future work, the method will cover more procedures from tobacco leaf grading system such as grading tobacco leaf based on leaf position and grading tobacco leaf based on a detailed color level.

Acknowledgments

The authors would like to thank to the Ministry of Research and Technology, the Republic of Indonesia for funding this research through the Higher Education Leading Development Research (Penelitian Pengembangan Unggulan Perguruan Tinggi – PPUPT) and PTPN X Klaten, Center Of Java as research partners.

Literature Cited

Bin, J., Ai, F.-F., Fan, W., Zhou, J.-H., Yun, Y.-H. and Liang, Y.-Z. 2016. A modified random forest approach to improve the multi-class classification performance of tobacco leaf grades coupled with NIR spectroscopy. *RSC Advances* 6(36): 30353–30361.

Chen, Y.-R., Chao, K. and Kim, M. S. 2002. Machine vision technology for agricultural applications. *Computers and Electronics in Agriculture* 36(2–3): 173–191.

Drope, J. and Schluger, N. W. (Eds), 2018. *The tobacco atlas*, 6th ed., American Cancer Society, Inc., GA.

FAO 2018a. Tobacco, available at: www.fao.org/land-water/databases-and-software/crop-information/tobacco/en/ (accessed September 6, 2018).

FAO 2018b. Tobacco commodity rankings, available at: www.fao.org/faostat/en/#rankings/countries_by_commodity (accessed September 6, 2018).

Jianqiang, Z., Weijuan, L., Huaihui, Z., Ying, H., Panpan, Y., Changyu, L., Yanmei, Y. and Ming, L. 2018. Automatic classification of tobacco leaves based on near-infrared spectroscopy and non-negative least squares. *Journal of Near Infrared Spectroscopy* 26(2): 101–105.

Laykin, S. S., Alchanatis, V. V., Fallik, E. E. and Edan, Y. Y. 2002. Image–processing algorithms for tomato classification. *Transactions of the ASAE* 45(3): 851–859.

Liu, J., Shen, J., Shen, Z. and Liu, R. 2012. Grading tobacco leaves based on image processing and generalized regression neural network. 2012 IEEE International Conference on Intelligent Control, Automatic Detection and High-End Equipment, pp. 89–93.

Mallikarjuna, P. B. and Guru, D. S. 2013. *Fusion of texture features and SBS method for classification of tobacco leaves for automatic harvesting*, Springer, New Delhi, pp. 115–126.

Mashithoh, E. 2013. Pengembangan Model Penentuan Kualitas Buah Berdasar Parameter Warna. Dissertation, Universitas Gadjah Mada, Yogyakarta.

Ni, L.-J., Zhang, L.-G., Xie, J. and Luo, J.-Q. 2009. Pattern recognition of Chinese flue-cured tobaccos by an improved and simplified K-nearest neighbor classification algorithm on near-infrared spectra. *Analytica Chimica Acta* 633(1): 43–50.

PTPN X. 2016. Research and Development division of PTPN X Klaten. Manual *Pelatihan Calon Grader.* PT, Indonesia.

Zhang, F. and Zhang, X. 2011. Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation. *Sensors* 11(3): 2369–2384.