Feasibility study of sugarcane stalks separation from trash using PCA based on color space of digital photos

IOP Conf. Series: Earth and Environmental Science **301** (2019) 012050 doi:10.1088/1755-1315/301/1/012050

Feasibility study of sugarcane stalks separation from trash using PCA based on color space of digital photos

Lalita Panduangnate[1], Khwantri Saengprachathanarug[1,2,*], Jetsada Posom[1,2], Arthit Phuphaphud[1], Chanreaksa Chea[1], and Eizo Taira[3]

[1] Curriculum of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand
[2] Applied Engineering for Important Crops of the North East Research Group, Khon Kaen University, Khon Kaen, Thailand
[3] Faculty of Agriculture, Ryukyus University, Japan

E-mail: khwantri@kku.ac.th

Abstract. Percentage of trash in the harvested sugarcane is one of the key quality indicators and needed to be examined since it affects sugar production efficiency. Currently, in Thailand, measurements of trash are done by randomly sampling by labors, which leads to delay and high labor cost this article aims to study the possibility of using the Principle Component Analysis (PCA) on the digital photos to detect harvested cane out of the trash. The component values of 5 color spaces (RGB, HSV, LAB, XYZ, and YIQ) of sugarcane stalks and trash images were obtained and then analyzed using PCA. The thresholding equation of each space was determined and applied for detecting sugarcane stalk pixel. The results revealed that the quality percentage (QP) of detection of all 5 color spaces was in a range from 82.27% to 96.19%, while those of LAB and RGB were the best at 96.06% and 94.90% respectively.

1. Introduction

In sugarcane mill, the cane quality is an important factor that related to the revenue. The cane had low quality extremely affected the income and sugar price. Trash in cane is main factor which leads to low quality.

The meaning of trash in cane is sugarcane leaf, rock, soil, and sand. It leads to the low efficiency of the squeezing process and low sugar yield. For example, the sugar yield can lose 0.64 - 0.91% if there is 1% of trash in sugarcane [1-2]. Refer to the report by Industrial Division Sugarcane Sugar and Downstream industries in 2017 [3], the average percentage of trash in Thailand sugar mills was 8.73%. It made the sugar mills lose the sugar yield approximately 6.76% or 3.8 billion USD in money.

In the present, the analysis of trash needed a lot of space and labors. Moreover, the labors required specific skills which have to pass a training course of trash analysis. This process had detail and complicated procedures. Thus, there is potential to get some mistake in the process. Furthermore, many labors needed for the sampling process with every 1 hour in whole day. All of these constraints are costly and time-consuming [4].

In recently, the image processing technique was applied to detect pest and weed in several plants [5-7]. Digital camera was used as a qualitative and quantitative detection tool for agricultural product. Koc-San et al. [8] applied the RGB camera to extract the citrus lemon trees from soil and weed in field, the result gave a quality percentage (QP) in a range of 69.57%–75.12%.and Yano et al. [9] separate the weeds from sugarcane leaves, this solution achieved an overall accuracy of 82 %. The results showed a good accuracy. So, there is a possibility to use RGB digital camera for the detection of harvested sugarcane from trash. However, some types of trash in cane have a similar characteristic with the sugarcane. So, only the RGB color might not enough for this study.

However, in the colors system, there is a way of converting color data from one color space to another through a mathematical transformation. To make certain calculations more convenient is color space. There are other color spaces obtained from the digital camera such as HSV, LAB, XYZ and YIQ. They were used for inspection in both agriculture and medicine [10-13]. These color spaces were investigated. The aim of this research was to investigate the possibility of using the color spaces including RGB, HSV, LAB, XYZ, and YIQ to detect the harvested sugarcane stalk from trash using principle component analysis (PCA) technique.

2. Materials and Methods

Figure 1 shows the flowchart of the experiment to detect the harvested sugarcane stalk using the color spaces.

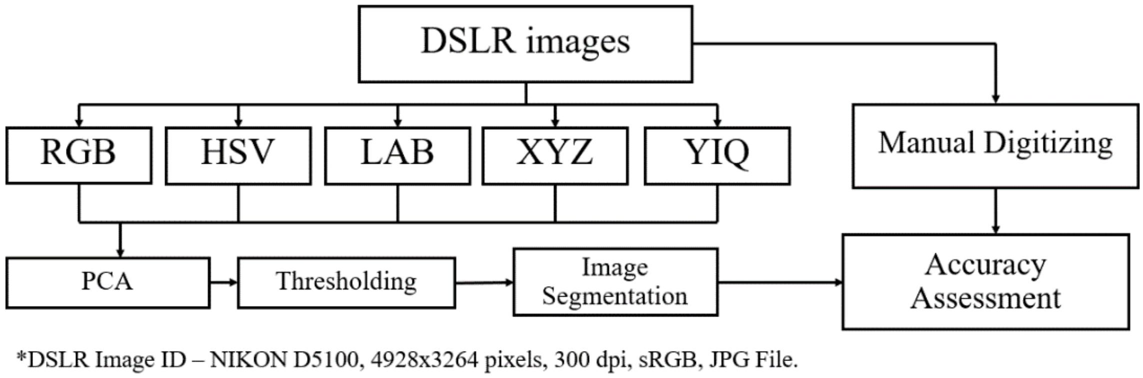

Figure 1. Schematic of the experiment

2.1. Material

Ten sugarcane stalks 'KK3 variety' at 13 months of age were collected. The leaves were removed and each cane stalk was cut into three parts (bottom, middle and top) in length of approximately 26-27 cm. Figure 2a shows the sample arrangement before picture acquisition in each sample. The material included background cleaned cane stalk and six main trash (fresh leaves, dry leaves, root, top leaves, soil and clay loaf), were applied for the image acquisition.

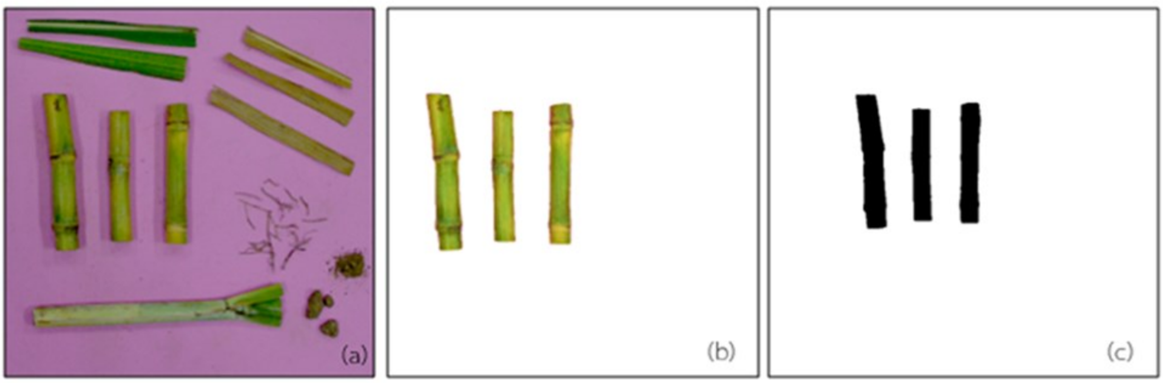

Figure 2. a) The original RGB picture of Sugarcane stalk arranged with six main trash (fresh leaves, dry leaves, root, top leaves, soil and clay loaf), (b) cropping sugarcane stalk image, and (c) a binary image of cropping cane stalk image.

2.2. Picture acquisition

Each cane stalk (bottom, middle and top parts) and cane trash (six types) was laid down on background (pink rubber foam) (see Figure 2a). The digital camera (D5100, Nikon, Thailand) was installed with a height of 90 cm above the ground and the light source was 4 lamp bulbs (T8, Philips, Thailand) placed in height of 2 meter above the ground. The cane stalk with trash was took a photo with the size of 4928 × 3264 pixels. The photography was done as above mention until ten sample.

Before analysis, all pictures were cropped into a size of 3000 × 3000 pixels. This step was done to get equal size (Figure 2a).

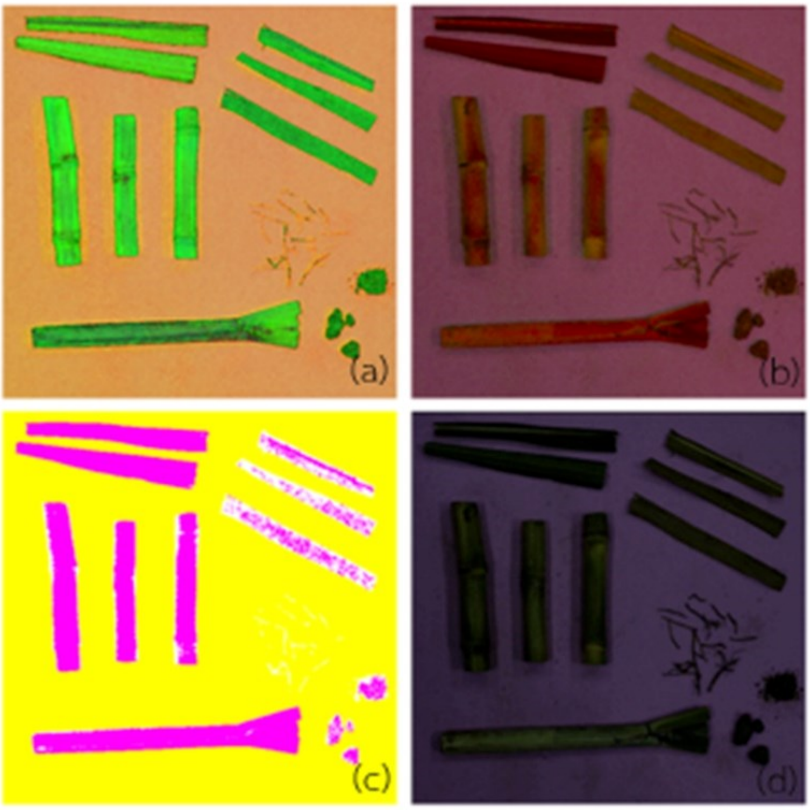

Figure 3. Color system in color spaces, (a) HSV, (b) YIQ, (c) LAB and (d) XYZ.

2.3. Sugarcane stalks detection

After collection, ten images were converted to color space types, i.e. HSV, XYZ, LAB and YIQ image with 32-bit component value, using Matlab R2018a (see Figure 3).

Five images were assigned as calibration set and five was used for the validation set. For calibration, each picture was cropped for the image size of 50 x 50 pixel by manually. The area of sugarcane stalk and six types of trash were selected and cropped randomly for five position.

To find the best color space to detect sugarcane stalk, the component value of each space were analyzed using PCA method as shown in Figure 4. The thresholding equations of each space represented as a solid line in Figure 4 was determined and applied to separate sugarcane stalk segmentation.

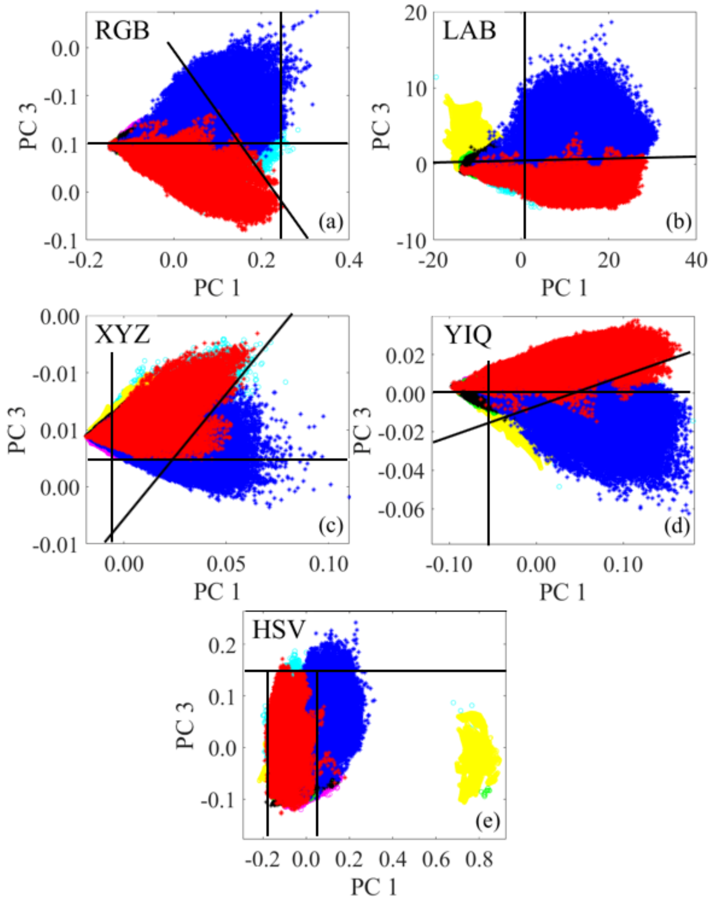

Figure 4. The PCA scores plot and line equation (solid line) for separation sugarcane stalks (Red Dot) in 5 color space.

2.4. Sugarcane stalks extraction

The purpose was to identify and count the number of sugarcane pixels. The interest image of only cane pixels was selected by visually and manually recognition using image editing software. Figure. 2b illustrated cropping sugarcane stalk image, then it was converted to the binary image. The pixels value in the binary image equal either 1 or 0 (represented in white and black color). The binary image was used as a standard method to count for the precision of image.

2.5. Accuracy assessment

In this study, the accuracy of cane stalk detection was assessed using quantitative assessment techniques. The separation between sugarcane and non-sugarcane pixels was done by counting the number of pixels after threshold process. The object extraction including True Positive (TP), False Negative (FN) and False Positive (FP) parameters were used to evaluate accuracy. The accuracy parameters of picture detection were Branching Factor (BF), Miss Factor (MF), Cane Detection Percentage (CDP) and Quality Percentage (QP), can be calculated as follows [14-15].

$$BF = (FP/TP) \quad (1)$$

$$MF = (FN/TP) \quad (2)$$

$$CDP = 100*TP/(TP+FN) \quad (3)$$

$$QP = 100*TP/(TP+FN+FP) \quad (4)$$

The ability of sugarcane detection between sugarcane sample and non-sugarcane was examined by BF, MF, CDP, and QP parameters. The performance in each color space was compared using one-way ANOVA. The difference in average of each color space was tested by the least significant difference (LSD) method.

3. Results and Discussion

3.1. Accuracy assessment results

For quantitative accuracy assessment, the reference and extracted sugarcane stalks were counted and analyzed using the component value of 5 color spaces. The results were shown in table 2. The results showed that the calibration set, and the validation set for all color space had the similar ability to detect sugarcane stalks. LAB was the color space that had the highest quality percentage (QP) of 96.19%, followed by RGB, XYZ, YIQ, and HSV which provided the QP of 94.62%, 89.34%, 88.46%, and 82.27% respectively. XYZ was a color space that was developed based on the basic theory of vision human eye which was sensitive to 3 colors of RGB. However, it still has limitations because the XYZ chart cannot describe Dark-Light characteristics of colors. Therefore, this color space has been developed to become a LAB color space for use as a device to reduce non-impartiality due to factors of origin and observers (International Commission on Illumination, CIE, Austria, 1931) that makes the LAB color space have capable of detecting color objects, while YIQ and HSV were developed for use in color systems of TV and computer graphics [16-17].

In Figure 5, showed the binary image of sample 1 created from thresholding with PCA. It was found that LAB image (Figure 5a) had a high ability to separate the trash of cane out of the image than that of HSV image (Figure 5b), resulting in higher QP while HSV image cannot recognize difference between trash and sugarcane, it becomes low QP.

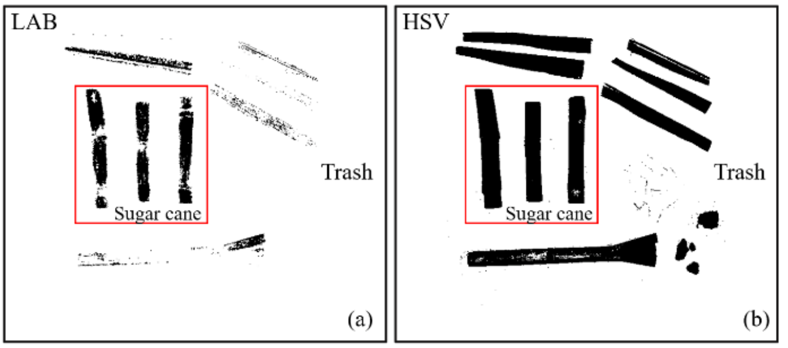

Figure. 5. The binary image sample 1 by PCA and threshold method for HSV (a) and LAB (b) Color Space.

When comparing the variance in evaluating the accuracy of all five color spaces using the One-way ANOVA analysis method, it was found that all 5 spatial parameters had a p-value <0.001, indicating that every color there is no difference statistically significant at 0.01, as shown in table 1.

Table 1. Results of ANOVA for 5 color spaces.

Data set	F-value	p-value
TP	38.165	0.0001**
FP	48.712	0.0001**
FN	69.832	0.0001**
BF	47.138	0.0001**
MF	44.438	0.0001**
CDP	47.325	0.0001**
QP	38.165	0.0001**

In the calibration set of table 2, shows average values of quantitative accuracy assessment parameters (TP, FP, FN, BF, MF, CDP, QP) of 5 color spaces (RGB, LAB, XYZ, YIQ, and HSV). The mean of each parameter of color space was compared by LSD. The results had statistically significant difference, at 1% significant level. The difference level can be divided into three groups. The first group was LAB and RGB, had no difference that gave the highest ability for detecting sugarcane stalks. The second group was XYZ and YIQ, and the last group occurring in HSV which has the lowest performance.

Table 2. The results of the average qualitative accuracy assessment of color space by PCA and thresholding method

Data set	Calibration set							Validation set						
	TP	FP	FN	BF	MF	CDP	QP	TP	FP	FN	BF	MF	CDP	QP
RGB	8,535,717 a	341,750 a	116,533 a	0.01 a	0.04 a	96.15 a	94.90 a	8,509,927	378,099	105,975	0.01	0.04	95.75	94.62
LAB	8,639,539 a	245,725 b	108,736 a	0.01 a	0.03 b	97.23 b	96.06 a	8,651,508	204,114	138,378	0.02	0.02	97.69	96.19
XYZ	8,215,625 a	278,688 ab	499,687 a	0.06 b	0.03 ab	96.72 ab	91.35 b	8,034,857	362,359	596,783	0.07	0.05	95.68	89.34
YIQ	8,040,756 b	64,600 c	888,642 c	0.11 d	0.01 c	99.20 c	89.40 b	7,956,012	74,661	963,327	0.12	0.01	99.07	88.46
HSV	7,382,134 c	48,057 c	1,563,810 b	0.21 c	0.01 c	99.36 c	82.08 c	7,399,066	55,662	1,539,273	0.21	0.01	99.25	82.27

TP, FP and FN (pixels), BF and MF (No unit), CDP and QP (Percentage). The same letter in a column are not significantly different at 1% level as determined by LSD.

4. Conclusions

This study showed the feasibility of using color spaces to detect the harvested sugarcane stalk from trash by PCA coupled with threshold techniques. Among the five color spaces, LAB provided the highest quality percentage (QP), while HSV gave the lowest QP. The color spaces of LAB, RGB, YIQ, XYZ, and HSV gave the QP of 96.19, 94.62, 89.34, 88.46, and 82.27%, respectively. There is the feasibility of detection the harvested sugarcane from trash using color spaces. However, color spaces gave low accuracy when applying with the unclean harvested sugarcane because their waxy color at sugarcane skin was similar to cane trash such as soil, dry leaves, and root. Thus, future studies should be generated using other wavelength bands or techniques such as near-infrared bands.

References

[1] Opanukul W, Nakwattananukul S, Jongsukwai K, Tunhaw M, Tongdaeng B, Sarntoonpituk D and Sookniyom S 2018 Study of Sugarcane Harvester Used in Thailand 13th TSAE *Natl. Conf*
[2] Werathavorn P 2003 Trash in cane is the farmer loss or profit? Chaiyaphum, Thailand: Mitr Phol Sugarcane Research Center. Ra-gubpit S. Thai Roong Ruang Group, the head of Thailand's largest sugar producer *Engineering today* pp 56-59
[3] Industrial Division Sugarcane Sugar and Downstream industries 2017 *The quantity of sugar produced in the factory*
[4] Office of the cane and sugar board 2017 *Production Efficiency of Sugar Production Season*, 134 pp 36-37
[5] Prabhu R, Prabhu G, Baskaran E and Arumugam E 2016 Clinical acceptability of metal-ceramic fixed partial dental prosthesis fabricated with direct metal laser sintering technique-5 year follow-up *J. Indian Prosthodont. Soc.* p 193
[6] Ding W and Taylor G 2016 Automatic moth detection from trap images for pest management *Comput. Electron. Agric.* pp 17–28
[7] Li Y, Xia C and Lee J 2015 Detection of small-sized insect pest in greenhouses based on multifractal analysis *Optik (Stuttg)* 126 pp 2138–43
[8] Koc-San D, Selim S, Aslan N and San B T 2018 Automatic citrus tree extraction from UAV images and digital surface models using circular Hough transform *Comput. Electron. Agric.* pp 289–301
[9] Yano I H, Alves J R, Santiago W E, and Mederos B J T 2016 Identification of weeds in sugarcane

fields through images taken by UAV and Random Forest classifier *IFAC-PapersOnLine* 49 pp 415–420
[10] Kolkur S, Kalbande D, Shimpi P, Bapat C and Jatakia J 2016 Human Skin Detection Using RGB, HSV and YCbCr Color Models *Proc. ICCASP* pp 324–332
[11] Prasetyo E, Adityo R. D, Suciati N and Fatichah C 2017 Mango leaf image segmentation on HSV and YCbCr color spaces using Otsu thresholding *ICST* pp 99–103
[12] Rahman M A, Edy Purnama I K and Purnomo M H 2015 Simple method of human skin detection using HSV and YCbCr color spaces *Proc. 2014 Int. Conf. Intell. Auton. Agents, Networks Syst. Ina.* pp 58–61
[13] Tran M T, Kim D H, Kim C K, Kim H K and Kim S B 2018 Determination of Injury Rate on Fish Surface Based on Fuzzy C-means Clustering Algorithm and L*a*b* Color Space Using ZED Stereo Camera *Int. Conf. Ubiquitous Robot* pp 466–471
[14] Larsen M, Eriksson M, Descombes X, Perrin G, Brandtberg T and Gougeon F.A 2011 Comparison of six individual tree crown detection algorithms evaluated under varying forest conditions *Int. J. Remote Sens*
[15] Yang B, Dai W, Dong Z and Liu Y 2016 Automatic forest mapping at individual tree levels from terrestrial laser scanning point clouds with a hierarchical minimum cut method *Remote Sens* pp 372
[16] Yachu H, Chulung C, Hsinyi C and Yiting T 2015 Application of YIQ Color Model for Color Pattern Recognition with Shifted and Rotated Training Images in Optoelectronic Correlator *Proc. Int. MultiConference Eng. Comput. Sci* pp 18–21
[17] Kondekar Vipul H, Bodhe S K 2018 A Comprehensive Investigation of Color Models used in Image Processing *IJCA* pp 19–24

Acknowledgments
The authors thank Sugar and Sugar Cane Research Center of the Northeast, Khon Kaen University and Applied Engineering for Important Crop of the Northeast Research Group, Khon Kaen University for support equipment, the test area and the facilities in this research.