

Robust weed recognition through color based image segmentation and convolution neural network based classification

M. Nazmuzzaman Khan¹

Mechanical and Energy Engineering Department
Indiana University-Purdue University Indianapolis
Indianapolis, Indiana

Sohel Anwar

Mechanical and Energy Engineering Department
Indiana University-Purdue University Indianapolis
Indianapolis, Indiana

ABSTRACT

Current image classification techniques for weed detection (classic vision techniques and deep-neural net) provide encouraging results under controlled environment. But most of the algorithms are not robust enough for real-world application. Different lighting conditions and shadows directly impact vegetation color. Varying outdoor lighting conditions create different colors, noise levels, contrast and brightness. High component of illumination causes sensor (industrial camera) saturation. As a result, threshold-based classification algorithms usually fail. To overcome this shortfall, we used visible spectral-index based segmentation to segment the weeds from background. Mean, variance, kurtosis, and skewness are calculated for each input image and image quality (good or bad) is determined. Bad quality image is converted to good-quality image using contrast limited adaptive histogram equalization (CLAHE) before segmentation. A convolution neural network (CNN) based classifier is then trained to classify three different types of weed (Ragweed, Pigweed and Cocklebur) common in a corn field. The main objective of this work is to construct a robust classifier, capable of classifying between three weed species in the presence of occlusion, noise, illumination variation, and motion blurring. Proposed histogram statistics-based image enhancement process solved weed mis-segmentation under extreme lighting condition. CNN based classifier shows accurate, robust classification under low-to-mid level motion blurring and various levels of noise.

Keywords: Image-segmentation, image-classification, precision-farming.

1. INTRODUCTION

With rapidly increasing global population, the demand for higher crop yield is also increasing rapidly. Weeds are one of the major culprits behind lower crop yield. They grow randomly in field and compete with crops for water, nutrients and sunlight. To eliminate weeds and reduce uncontrolled spray of herbicides, real-time detection of weeds with high accuracy using low cost sensors is needed. Among all the sensing techniques (machine vision, spectroscopy, fluorescence, LiDAR and ultrasonics) [1], this work will focus on RGB camera based machine vision.

CNN based weed classification gained popularity in recent times due to their generalization capability and hardware acceleration. Dyrmann et al. [2] achieved 86.2% accuracy with 22 species and 10413 images with a CNN network build from scratch. McCool et al. [3] achieved greater than 90% accuracy with 1.07-1.82 frame per second using a combination of lightweight CNN models. But classification accuracy of CNN models are highly dependent on motion blur, noise, and contrast [4]. Any work on the effect of noise, blur and contrast for real-time weed detection, and steps to mitigate them have not been well investigated.

Any real-time robust classification method is based on redundant system. In this case, we deploy visible spectral-index based image segmentation (weed segment from background) as fail-safe when CNN performance deteriorates. Both the segmented image of the weed and classified image of the weed can be feed into a decision-making system. Based on the performance of the CNN, the decision-making system can decide which herbicide to spray. The goal is to not miss any weed (or not spray wrong herbicide) even if CNN misclassified. The block diagram is showed in Fig 1. ExG [5], ExGR [6] and CIVE [7] are color based segmentation methods and they all perform badly when light is high or low. Guijarro et al. showed that

¹ Contact author: mdkhan@iu.edu

combination of different color-based segmentation methods can be ensemble to produce better segmentation result [8]. But they still poorly perform under extreme lighting condition. Pre-processing is needed to control the lighting of the image before using it for segmentation.

In this study, we have proposed a histogram-based image statistics system to identify image quality. If image is low-quality, it passes through CLAHE step for histogram equalization. Eventually good-quality image is passed to visible spectral based segmentation step for weed segmentation from background. Same good-quality image is passed through transfer-learning based CNN classifier for weed classification. Both the segmented weed image and classified weed image is passed to a decision-making system. Based on illumination, noise and motion blur, CNN based weed classification accuracy can vary. Visible spectral based weed segmentation is a fail-safe to the CNN classifier, so that spraying herbicide on weed is not missed due to CNN misclassification.

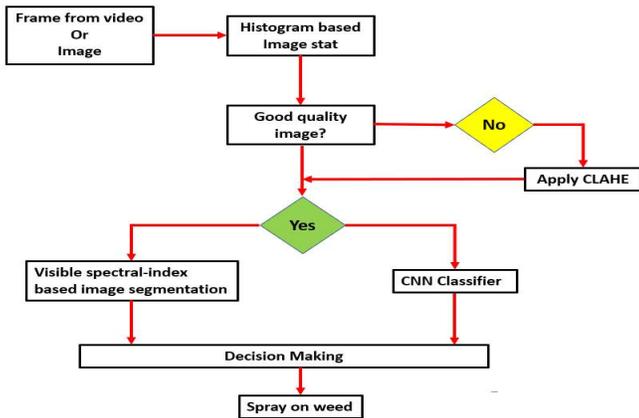


Figure 1: Block diagram of proposed image enhancement process for precision farming.

2. COLOR BASED IMAGE SEGMENTATION

Given an input image in RGB color space, each channel is separated by following system:

R = red channel of input image

G = green channel of input image

B = blue channel of input image

After splitting the channels, the following normalization scheme is applied, which is common in agronomic image segmentation [9]:

$$r = \frac{R_n}{R_n + G_n + B_n}, g = \frac{G_n}{R_n + G_n + B_n}, b = \frac{B_n}{R_n + G_n + B_n} \quad (1)$$

where R, G and B are the normalized RGB coordinates ranging from 0 to 1 and are obtained as follows:

$$R_n = \frac{R}{R_{max}}, G_n = \frac{G}{G_{max}}, B_n = \frac{B}{B_{max}} \quad (2)$$

where R_{max} , G_{max} and B_{max} are maximum values of R, G and B channels respectively. A small number is added to the denominator of normalization step to avoid division by zero.

Green color (vegetation) can be extracted using the following equations:

$$\text{Excess green [10]: } \text{ExG} = 2g - r - b \quad (3)$$

$$\text{Excess green minus excess red [6]: } \text{ExGR} = \text{ExG} - 1.4r - g \quad (4)$$

Color index of vegetation extraction [7]:

$$\text{CIVE} = -(0.441r - 0.811g + 0.385b + 18.78745) \quad (5)$$

The above three methods are combined to get the resulting green color segmentation:

$$\text{GREEN} = w_{\text{ExG}} * \text{ExG} + w_{\text{ExGR}} * \text{ExGR} + w_{\text{CIVE}} * \text{CIVE} \quad (6)$$

where w_{ExG} , w_{ExGR} and w_{CIVE} are weights for each index, representing the relative importance of the index. The following weight values are used in this study: $w_{\text{ExG}} = 0.28$, $w_{\text{ExGR}} = 0.34$, $w_{\text{CIVE}} = 0.38$. The resulting combined image (GREEN), is linearly mapped to range in [0, 255], after which, it is thresholded by applying the Otsu's [11] method, obtaining a binary (single channel) image. Otsu's method assumes bi-modal distribution of histogram and calculates the value between two histogram peaks. In this study, we are segmenting weed from background (foreground-background segmentation) which is bi-modal. So, using Otsu's method is justifiable. The binary image then converted back to RGB (3 channel) image. Here green pixels identify plants in the original image with clear spectral RGB components and white pixels identify soil and other materials belonging to the background. This method is designed to cope with the variability of natural daylight illumination. Fig 2 shows how image is segmented using 4 different methods.



Figure 2: ExG, ExGR, CIVE and GREEN based image segmentation. GREEN is able to capture most of the weeds using combination of other three methods.

Fig. 3 shows performance of different segmentation method under dark and bright lighting conditions. Under dark condition, ExGR performed better and was able to segment the weed properly. GREEN was able to capture most of the plant. Under bright light, ExGR failed completely to segment the weed from background. ExG and CIVE was able to segment most of the weed from background. This implicates two things. Firstly, although visible spectral-index based image segmentation was designed to cope with the variability of natural daylight

illumination, they may fail under extreme condition. As a result, combination of different methods will be more robust and will be better at segmentation under different condition. Secondly, to prevent segmentation from failing, we have to understand when it is happening. We have to quantitatively measure the difference between ‘good-quality’ and ‘bad-quality’ image and convert the ‘bad-quality’ image into ‘good-quality’ image. That’s where the histogram-based image statistics comes in.

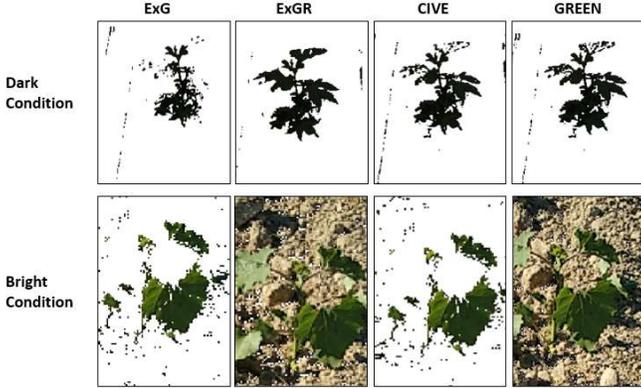


Figure 3: ExG, ExGR, CIVE and GREEN based image segmentation performance under different lighting conditions.

3. HISTOGRAM BASED IMAGE STATISTICS:

A histogram represents the distribution of pixel intensities (color or gray scale) in an image. When plotting the histogram, the X-axis serves as pixel values. For RGB color space, each channel will have pixel values in the range of 0 to 255 (for 8-bit image). The Y-axis indicates number of pixels. When we construct a histogram with 0-255 pixel values (bins), then we are effectively counting the number of times each pixel value occurs at a certain bin. If number of pixels (Y-axis) is high at bin (X-axis) value 0, the image is generally darker. If number of pixels (Y-axis) is high at bin (X-axis) value 255, the image is generally brighter. Examining the histogram of an image, a general idea of the contrast, brightness and intensity distribution can be achieved.

For gray image (single channel), histogram is defined by discrete function $h(a) = n_a$, where ‘a’ represents gray level (X-axis) and n_a represents number of pixels (Y-axis) for each value of ‘a’. Probability of each ‘a’, $p(a) = h(a)/(M * N)$, where M and N are rows and columns of the image respectively. Let ‘a’ be a random variable denoting gray levels, the nth moment of ‘a’ about the mean is defined as [12]:

$$\mu_n = \sum_a (n(a) - m)^n p(a) \quad (7)$$

Where m is mean value of sum of $n(a)$

$$m = \sum_a n(a) \cdot p(a) \quad (8)$$

Following additional statistical parameters are calculated:

$$\text{variance} = \mu_2(a), \text{skewness} = \frac{\mu_3(a)}{\mu_2^{1.5}(a)} \quad (9)$$

$$\text{kurtosis} = \frac{\mu_4(a)}{\mu_2^2(a)}$$

Following information can be summarized from the statistical parameters [5]. The mean determines the average level of brightness. Variance is a measure of gray-level contrast, where high values indicate dispersion of values around the mean and low values are indicative of a high concentration of values around the mean. The skewness measures the asymmetry in the distribution. A positive skewness is presented when the histogram has low values around high brightness (around X-axis pixel value 255) and high values in the part of low brightness values (around X-axis pixel value 0). In the opposite case the skewness is negative. The kurtosis provides information about how the distribution behaves around the peak. Low kurtosis indicates flat top parts in the histogram around the mean but high values are indicative of peaks around the mean with high slopes. Skewness and kurtosis are both zero for Gaussian distributions.

According to Romeo et al. [7], images highly contrasted are considered as images with sufficient quality and vice versa. An image with sufficient contrast should be identified by mean values in the central part of histogram, high variance, low skewness (positive or negative) and high kurtosis. On the contrary, an image with insufficient contrast is identified by mean values either low or high, high skewness (positive or negative) and low kurtosis. The next step is to determine the ranges of variability for the above parameters. The goal is to use these parameters to convert low quality images into high quality images.

4. PROPOSED BRIGHTNESS/CONTRAST CONTROL STEPS:

The shortcoming of histogram enhancement (HE) is over-enhancement in images with large smooth area. For images captured under low light condition, HE causes over-enhancement after contrast enhancement and increases the noise. Contrast limited adaptive histogram equalization (CLAHE) overcomes the over-enhancement problem of HE by minimizing noise-like artifacts in homogeneous regions [13]. As we have seen already, visible spectral-index based image segmentation fails for bad-quality image. Applying CLAHE to bad-quality image before segmentation will solve this problem. In this study we propose an automatic method to define image quality based on image histogram statistics so that low-quality images can be converted to high quality images. We have tested the effects of CLAHE clip-limit and step-size on histogram enhancement and clip-limit = 3.0 and grid-size = (8x8) is fixed for this application. Steps are as follows:

(step 1) Determine the ranges of good quality image parameters (mean, variance, skewness and kurtosis) using a set of training images. See equation (10) for parameters.

(step 2) Calculate mean, variance, skewness and kurtosis of test image for R, G and B channels. In this case only mean and kurtosis according to equation (10). Determine if good or bad quality image.

(step 3) For bad-quality image, apply CLAHE to convert it to good-quality image.

(step 4) Apply visible spectral-index based image segmentation (GREEN) on good-quality image.
 (step 5) Apply morphological operation (opening and closing) on segmented image to fill-up void and reduce noise.

Fig 4 and 5 shows example of good-quality and bad-quality images respectively. Good-quality image has mean at the center (relatively), higher variance and lower skew and kurtosis compared to bad quality image. Fig 6 shows how CLAHE affects the histogram. Original image failed GREEN segmentation because blue-channel is left tilted (low-quality image). CLAHE equalized the histogram and successfully segmented the image. CLAHE step needed 0.212 sec to process a 150x150x3 image.

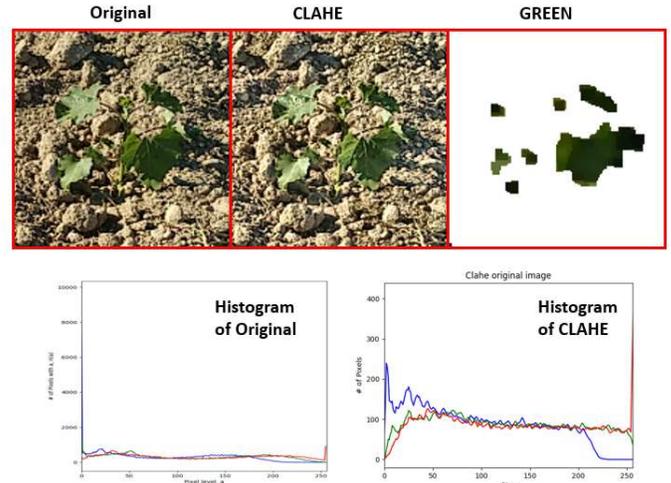


Figure 6: Original bad-quality image, bad-quality image after CLAHE, visible spectral-index based image segmentation (GREEN) on CLAHE applied image, histogram of original image, histogram of CLAHE image.

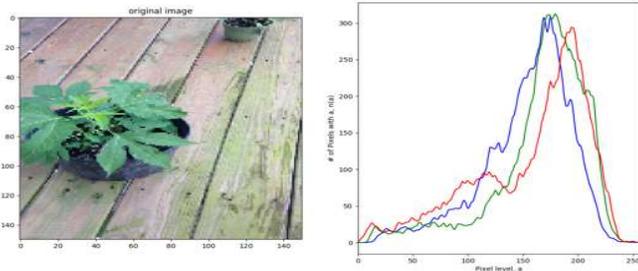
4.1 Image quality: Identification of parameters

Now that we have an idea how good-quality and bad-quality images affect the histogram-based statistics parameters, we need to determine the ranges of these parameters. When proper ranges of good-quality image parameters are determined, CLAHE can be applied on bad-quality images to convert them into good-quality images. Table 1 and 2 shows mean, variance, skew and kurtosis (statistics parameters) values in R, G, B channels for good-quality and bad-quality images respectively. 15 images were selected for both (good and bad) cases and their minimum, maximum and average values are shown in Table 1 and 2.

Table 1: Mean, variance, skewness, kurtosis values for good-quality images in the three R, G, B spectral channels.

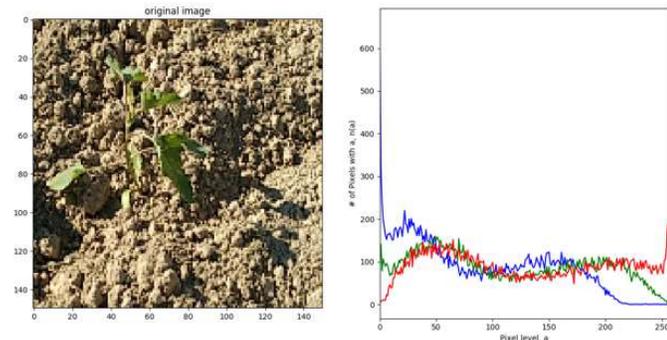
		R	G	B
Mean	Max	663.54	411.10	365.56
	Min	157.54	170.80	153.39
	Avg	288.85	253.64	243.2
Variance	Max	246758.1	96448.55	68238.15
	Min	7472.22	7577.26	5685.19
	Avg	54515.4	30349.61	24210.3
Skew	Max	0.37	0.62	0.344
	Min	0.009	0.021	0.013
	Avg	0.19	0.29	0.166
Kurtosis	Max	1.73	2.07	1.98
	Min	1.47	1.50	1.52
	Avg	1.64	1.77	1.69

Table 2: Mean, variance, skewness, kurtosis values for bad-quality images in the three R, G, B spectral channels.



	R	G	B
Mean	157.5	200.63	183.6
Variance	7472.2	9181.6	7581.7
Skew	0.124	0.62	0.34
Kurtosis	1.59	2.07	1.98

Figure 4: A good-quality image with histogram and statistics parameters for R, G and B channels of the image.



	R	G	B
Mean	96.44	99.18	140.43
Variance	646.96	841.81	10778.7
Skew	0.35	0.102	3.70
Kurtosis	3.45	2.43	18.80

Figure 5: A bad-quality image with histogram and statistics parameters for R, G and B channels of the image.

		R	G	B
Mean	Max	391.1	368.50	436.65
	Min	96.4	99.18	140.43
	Avg	198.4	192.39	235.39
Variance	Max	33391.47	35723.3	53021.5
	Min	646.9	841.81	9955.4
	Avg	11090.1	11355.63	22818.8
Skew	Max	1.15	0.87	3.72
	Min	0.35	0.10	0.32
	Avg	0.76	0.54	2.06
Kurtosis	Max	5.13	4.6	18.8
	Min	2.06	1.88	1.87
	Avg	3.27	2.78	10.13

From Table 1 and 2, several observations can be made. In general, good quality image (Table 1) has mean at the center (relatively), higher variance and lower skew and kurtosis compared to bad quality (Table 2) image. For all three channels, parameter values overlap for mean, variance and skew; for good and bad quality images. Bad quality maximum mean value is higher than good quality minimum mean value. Same is also true for variance and skew. For kurtosis, overlapping occurs for G and B channels. This indicates that the training weed dataset available to the authors are not diverse enough to clearly distinguish between good quality and bad quality images. But from Table 1, statistics features for good quality images are clearly identified. Skew and kurtosis parameters are chosen because they are reasonably distinguishable between good and bad quality images. This step needed on an average 0.0023 sec to process a 150x150x3 image. The following criterias are chosen to distinguish the bad quality image from good quality image. Good quality image criteria are:

$$\begin{aligned}
 R \text{ channel kurtosis} &< 1.73 \\
 G \text{ channel kurtosis} &< 2.07 \\
 B \text{ channel kurtosis} &< 1.98 \\
 R \text{ channel skew} &< 0.37 \\
 G \text{ channel skew} &< 0.62 \\
 B \text{ channel skew} &< 0.34
 \end{aligned} \tag{10}$$

In conclusion, for every image kurtosis and skew will be calculated. If skew and kurtosis are within range of Eq. (10), it is good quality image and GREEN is applied for segmentation (Eq. (6)). If bad quality, CLAHE will be applied to make the image good quality. Then segmentation will be applied.

5. TRANSFER-LEARNING BASED CNN CLASSIFIER:

An image convolution is an element-wise multiplication of two matrices followed by a sum. An image is essentially a multidimensional matrix. It has width (number of columns), height (number of rows) and depth (number of channels – for RGB image it's 3). This big matrix (image) is multiplied with a

small matrix (kernel) to create the convolution operation. In specific computer vision application (like edge detection), kernel is hand-defined. As an example, for Sobel edge detection, kernel is a 3*3 matrix with zero values at the center column. A machine learning algorithm designed to look at the training images and create these kernels (or filters) to detect specific objects are called convolution neural network. A convolution works by sliding these windows of size 3×3 or 5×5 ($n \times n$ sized kernels) over the 3D input feature map (image), stopping at every possible location, and extracting the 3D patch of surrounding features.

Table 3: Weed image dataset

Dataset	Number of images		
	Cocklebur	Pigweed	Ragweed
Training (used for training the classifiers)	544	505	552
Validation (used for tuning hyperparameters during training)	65	62	69
Testing (used for confusion matrix – never seen by the classifier)	65	62	69

Transfer learning retrains the final layer of the VGG model to classify a new dataset by exploiting the large amount of visual knowledge already learned from the Imagenet database. VGG16 is trained from the ImageNet Large Visual Recognition Challenge using the data from 2012, where it was tasked with classifying images into 1,000 classes. The top-5 error rate of VGG16 was 7.4% [14]. The model parameters implemented in this study included the steps per epoch=100, epochs=50, learning rate (1e-05), train batch size (20), and the validation batch size (20), optimizer (rmsprop), loss (sparse categorical cross-entropy). On a core-i5 8gb ram machine, for input image size 150 x 150 x 3, it took 0.266 secs to process and classify. Table 3 shows the weed dataset size used for this study. Accuracy plot and confusion matrix of the CNN model is presented in Fig 7 and 8 respectively.

5.1 Effect of noise and motion blur on CNN classifier:

Brightness of the image is already controlled by image quality identification step. As a result, effect of noise and motion blur on classification accuracy is evaluated in this step. Image capturing from a vehicle on a corn field may produce blurriness, which will affect the classification accuracy. 3 stages of motion blurriness (low (10%), medium (25%), high (40%)) is artificially induced on 2 sets (near and far field) of images (Fig 9). For low-blur, both near and far field image is correctly classified. For near-field, incorrect classification occurs at medium and high blur. For far-field, incorrect classification occurs at only high-blur. Motion blur has higher impact on near-field image classification inaccuracy than far-field image. This led to the

conclusion that, camera should be set at a distance which correctly captures the features of the weed plants from as far as possible. Also, traditional masks (Unsharp Mask and Gaussian Mask) failed to correct the motion blurriness at high-blur. Which indicates, when information is lost due to motion blur, it is hard to recapture.

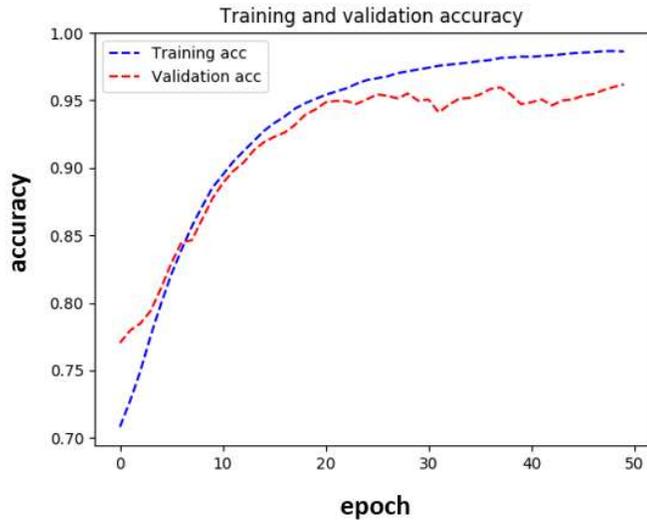


Figure 7: Training and validation accuracy of CNN based classifier.

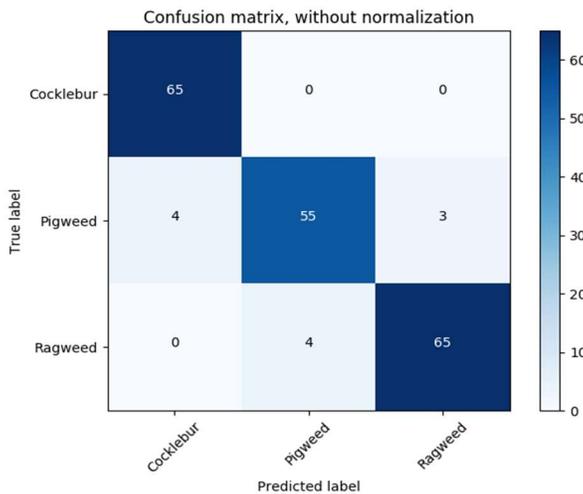


Figure 8: Confusion matrix of the CNN based classifier.

Field condition and low-quality sensor can cause noise in images. Noise can affect the accuracy of a classifier. In this case, gaussian and salt-and-pepper noise is artificially introduced in images to evaluate classifier performance. As seen from Fig 10, noise doesn't affect the classifier accuracy. Varying the intensities also didn't affect the classification.

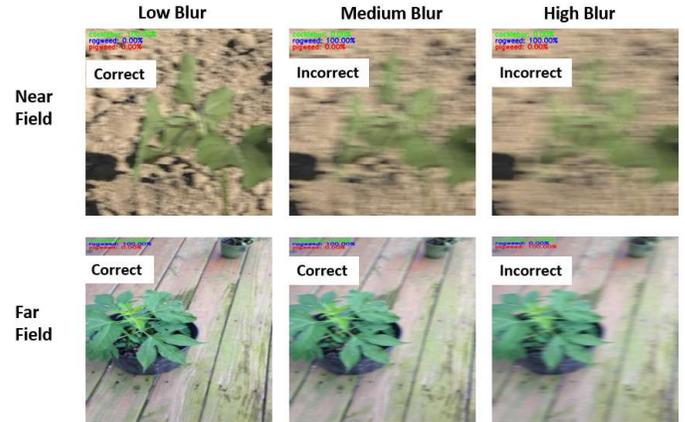


Figure 9: Effect of 3 stages of motion blur (low, medium, high) on classification accuracy. Top row shows the effect on near-field image and bottom row on far-field image.

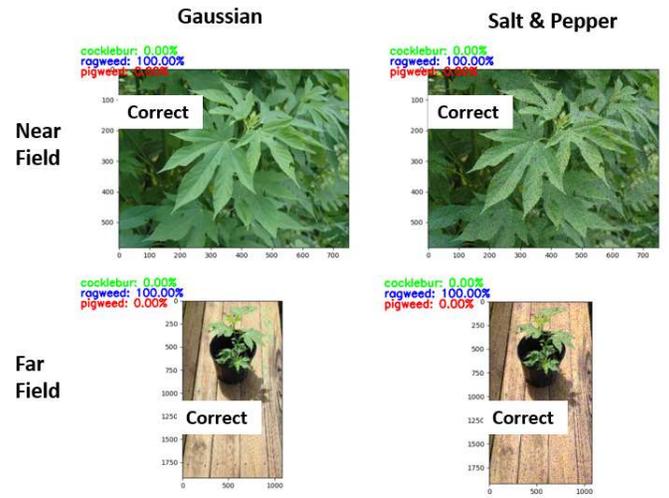


Figure 10: Effect of Gaussian and salt-and-pepper noise effect on classification accuracy. Top row shows the effect on near-field image and bottom row on far-field image.

6. CONCLUSION:

The proposed algorithm using histogram-based image statistics for image quality determination and brightness control significantly improved the inherent limitation of color-based segmentation (bad performance under extreme lighting) under various test conditions. The real-time processing of the histogram based statistical algorithm was only 0.0023 seconds which is significantly shorter than that of classification (0.266 secs) and CLAHE (0.212 secs) methods. Such a fast processing time of the proposed histogram based algorithm is suitable for real-time control applications. It is also observed that CNN based classifier performed better for far-field images because the near-field images lost more information for the same amount of blur. Classification accuracy for this method was unaffected by Gaussian and salt-and-pepper noises. When both the histogram based algorithm and CNN are combined for image classification for weed detection, the overall the system exhibited robust

performance in classification under varying lighting, blur, and noise conditions.

REFERENCE

1. Wang, A., W. Zhang, and X. Wei, *A review on weed detection using ground-based machine vision and image processing techniques*. Computers and Electronics in Agriculture, 2019. **158**: p. 226-240.
2. Dyrmann, M., H. Karstoft, and H.S. Midtby, *Plant species classification using deep convolutional neural network*. Biosystems Engineering, 2016. **151**: p. 72-80.
3. McCool, C., T. Perez, and B. Upcroft, *Mixtures of lightweight deep convolutional neural networks: applied to agricultural robotics*. IEEE Robotics and Automation Letters, 2017. **2**(3): p. 1344-1351.
4. Dodge, S. and L. Karam. *Understanding how image quality affects deep neural networks*. in *2016 eighth international conference on quality of multimedia experience (QoMEX)*. 2016. IEEE.
5. Romeo, J., et al., *A new Expert System for greenness identification in agricultural images*. Expert Systems with Applications, 2013. **40**(6): p. 2275-2286.
6. Camargo, N.J., *A combined statistical-soft computing approach for classification and mapping weed species in minimum-tillage systems*. 2004.
7. Kataoka, T., et al. *Crop growth estimation system using machine vision*. in *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*. 2003. IEEE.
8. Guijarro, M., et al., *Automatic segmentation of relevant textures in agricultural images*. Computers and Electronics in Agriculture, 2011. **75**(1): p. 75-83.
9. Gée, C., et al., *Crop/weed discrimination in perspective agronomic images*. Computers and Electronics in Agriculture, 2008. **60**(1): p. 49-59.
10. Ribeiro, A., et al., *Development of an image analysis system for estimation of weed pressure*. Precision agriculture, 2005. **5**: p. 169-174.
11. Otsu, N., *A threshold selection method from gray-level histograms*. IEEE transactions on systems, man, and cybernetics, 1979. **9**(1): p. 62-66.
12. Gonzalez, R. and R. Woods, *Digital image processing: Pearson prentice hall*. Upper Saddle River, NJ, 2008. **1**.
13. Chang, Y., et al., *Automatic contrast-limited adaptive histogram equalization with dual gamma correction*. IEEE Access, 2018. **6**: p. 11782-11792.
14. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.