Yu et al. Plant Methods (2024) 20:155
https://doi.org/10.1186/s13007-024-01281-5

Plant Methods

RESEARCH

Study on canopy extraction method for narrowband spectral images based on superpixel color gradation skewness distribution features

Hongfeng Yu[1], Yongqian Ding[2,3*], Pei Zhang[4], Furui Zhang[2], Xianglin Dou[1] and Zhengmeng Chen[5]

Abstract

Background Crop phenotype extraction devices based on multiband narrowband spectral images can effectively detect the physiological and biochemical parameters of crops, which plays a positive role in guiding the development of precision agriculture. Although the narrowband spectral image canopy extraction method is a fundamental algorithm for the development of crop phenotype extraction devices, developing a highly real-time and embedded integrated narrowband spectral image canopy extraction method remains challenging owing to the small difference between the narrowband spectral image canopy and background.

Methods This study identified and validated the skewed distribution of leaf color gradation in narrowband spectral images. By introducing kurtosis and skewness feature parameters, a canopy extraction method based on a superpixel skewed color gradation distribution was proposed for narrowband spectral images. In addition, different types of parameter combinations were input to construct two classifier models, and the contribution of the skewed distribution feature parameters to the proposed canopy extraction method was evaluated to confirm the effectiveness of introducing skewed leaf color skewed distribution features.

Results Leaf color gradient skewness verification was conducted on 4200 superpixels of different sizes, and 4190 superpixels conformed to the skewness distribution. The intersection over union (IoU) between the soil background and canopy of the expanded leaf color skewed distribution feature parameters was 90.41%, whereas that of the traditional Otsu segmentation algorithm was 77.95%. The canopy extraction method used in this study performed significantly better than the traditional threshold segmentation method, using the same training set, Y1 (without skewed parameters) and Y2 (with skewed parameters) Bayesian classifier models were constructed. After evaluating the segmentation effect of introducing skewed parameters, the average classification accuracies Acc_Y1 of the Y1 model and Acc_Y2 of the Y2 model were 72.02% and 91.76%, respectively, under the same test conditions. This indicates that introducing leaf color gradient skewed parameters can significantly improve the accuracy of Bayesian classifiers for narrowband spectral images of the canopy and soil background.

*Correspondence:
Yongqian Ding
yongqiand@njau.edu.cn

Full list of author information is available at the end of the article

Conclusions The introduction of kurtosis and skewness as leaf color skewness feature parameters can expand the expression of leaf color information in narrowband spectral images. The narrowband spectral image canopy extraction method based on superpixel color skewness distribution features can effectively segment the canopy and soil background in narrowband spectral images, thereby providing a new solution for crop canopy phenotype feature extraction.

Keywords Narrowband spectral image, Superpixel, Skewed distribution, Canopy extraction, Color gradation

Background

In modern agriculture, researchers have developed various types of crop phenotype extraction devices using qualitative and quantitative relationships between crop physiological and biochemical parameters and specific narrowband spectral images [1–3]. The most commonly used devices are multispectral drones (such as the DJI Multispectral serial products), which have several built-in narrowband spectral cameras that can obtain real-time narrowband spectral phenotypes of crops in the field [4, 5]. Researchers can monitor the physiological and biochemical status of crops using these phenotypes as a basis for precision agricultural management [6, 7].

The analysis accuracy of narrowband spectral images in specific bands may be affected by background impurities such as soil and straw. In particular, when the crop canopy coverage is low, soil background interference has a significant impact. If the necessary soil and canopy signal separation processing is not performed, the soil background will seriously affect the accuracy of crop phenotype extraction [8, 9]. However, narrowband spectral images differ from conventional RGB images in that the difference between the canopy and background is small, making it difficult to apply existing RGB image background segmentation denoising methods (such as traditional threshold segmentation [10] and color segmentation GMR [11]) for accurate separation. Scholars have explored this using RGB cameras arranged side by side and binocular matching algorithms to perform intersection detection on narrowband spectral camera channel image data, achieving foreground and background separation of the narrowband spectral images [12]. Although the separation accuracy has been improved, this method has a complex technical implementation and high equipment costs, and is not universal, making it unsuitable for practical production. Therefore, the precise extraction of the crop canopy based on narrowband spectral images remains challenging.

With the rapid development of artificial intelligence, machine learning has been widely used to separate and extract target objects in RGB digital images. Feature parameter decision-making and convolutional neural networks are two common technical methods of machine learning [13, 14]. Owing to the large number of training samples required for convolutional neural network segmentation, high hardware requirements during training, and the overfitting phenomenon, especially in the face of complex and variable lighting environments in the field, the generalization and adaptability of such models are poor [15, 16]. Therefore, in the field of RGB image phenotype separation and extraction, researchers often use feature parameter decision-making methods with small training volumes and high processing speeds. However, this method is limited by the dimensions of the feature parameters of the image to be detected; in cases where the parameter dimension is small, the separation and extraction accuracy cannot meet the application requirements [17].

At present, feature parameter decision-making methods mostlyprimarily deal with superpixels, which consider the adjacency of pixels and divide the image into several sub regionssubregions with semantic significance. This is more conducive to the extraction of local features and the expression of structural information, greatly reducing the computational complexity of the subsequent processing [18, 19]. Therefore, some scholars have used this technology to improve the segmentation accuracy of feature parameter decision- making methods. For example, Lu et al. [20] fused the simple linear iterative clustering (SLIC) method to generate superpixels, and imported the mean color scale of the superpixels into a neural network population model. Corn ears were segmented by distinguishing the types of superpixels. Yang et al. [21] used the SILC algorithm to generate citrus canopy superpixels and constructed a BPNN classifier based on the color and texture parameters of the superpixels to extract mature citrus fruits. However, the above studies treated the color distribution of superpixels as a normal distribution or an approximate normal distribution [22], which results in these methods obtaining only one phenotype feature parameter, the mean of each channel, and less information. This can only be used as a basis and HSV, LAB, and other parameters can be combined to form a composite parameter system to describe crop canopy color features, thereby greatly limiting the feature expression of superpixels and the classification accuracy of the constructed classifier [23].

With in-depth research on the RGB color space of crop canopies, researchers have demonstrated the existence of a skewed distribution of canopy color gradation in various crops, such as soybeans, cabbage, chili peppers, purple-backed sky sunflowers, and tobacco [24–26]. On this basis, 20 commonly used RGB model skewness parameters, including the mean, median, mode, skewness, and kurtosis, have been obtained, expanding the RGB color feature parameter system and obtaining a more effective multiclass feature parameter decision model. Taking inspiration from the above, this study focuses on narrowband wheat spectral images in the 770 nm wavelength range with different illuminances. The extraction and application of superpixel skewness distribution parameters are used as innovative means, and the superpixel skewness distribution parameters of the narrowband spectral images are used as inputs to construct a Bayesian classifier for the canopy and soil background of the narrowband spectral images. In this study, the problem of narrowband spectral image canopy segmentation and extraction is transformed into the problem of superpixel classification, and the performance of this method is compared with that of the traditional threshold segmentation method to clarify its progressiveness and universality. In addition, we compare the classification accuracy of Bayesian models under different parameter systems to elucidate the role and significance of skewed distribution parameters in this method further.

Materials and methods

Test conditions and image acquisition

The experimental site is located at the Baima Base of Nanjing Agricultural University, Nanjing, Jiangsu Province, China. Different wheat varieties were planted in experimental plots, with a single plot area of approximately 1 m × 0.8 m. Four rows of wheat were planted with a row spacing of 25 cm. The experimental setup for the test locations is shown in Fig. 1. The narrowband spectral camera was fixed to a movable bracket at a height of 2 m. The narrowband spectral camera model was a full-band spectral WP-UC600 industrial camera for imaging with a built-in narrowband filter center wavelength of 770 nm and bandwidth of 20 nm. The narrowband spectral image pixels were 3072×2048 with a frame rate of 30 fps. The field-of-view angle of the camera was 57.6° × 49.3° × 71.2° and the camera exposure time was fixed at 30 ms. Each wheat plot image was captured at a fixed point using a movable bracket.

In general, data were collected during the unsealed stage of wheat, with 12 varieties and 24 unit images. The specific collection dates were February 21, 2021, and March 2, 2021, with a total of 48 images collected. The sampling time interval for the experiment was from 08:00 to 16:00, and the range of the collected solar irradiance variation was approximately 5000 lx to 10 0000 lx. Meanwhile, all obtained images were uniformly adjusted in Python 3.8 to generate JPG format files, following which feature parameter extraction and canopy extraction were performed on the images using Python 3.8. An NVIDIA Jetson TX2 processor was used.

Canopy extraction method for narrowband spectral images based on superpixel color gradation skewness distribution features

Extraction process

The entire processing flow of the narrowband spectral image canopy extraction method, based on the skewed

Fig. 1 General image of test site

distribution characteristics of the superpixel color gradation, is shown in Fig. 2. Step 2 (superpixel decomposition) mainly functions to decompose the narrowband spectral image into superpixels for classification, and the number of superpixel decompositions K and the tightness L need to be specified. Step 3 (narrowband spectral image feature parameter extraction) is the core of this study, in which a system of skewed distribution parameters is introduced for superpixel scale narrowband spectral images. Step 4 (training and prediction of Bayesian classifier) takes the input as the feature parameters of Step 3 and constructs a Bayesian classifier model on this basis to achieve the classification of narrowband spectral image canopy and soil background superpixels. Step 5 (canopy superpixel stitching and denoising) refines the extracted canopy boundaries to improve the separation quality. These steps are described in detail in the following sections.

Superpixel decomposition

The currently widely used and effective superpixel algorithm is SLIC [27, 28]. This algorithm iteratively clusters image pixels based on the color similarity and spatial distance relationships. The SLIC algorithm constructs a five-dimensional feature vector K (where K represents the total number of image pixels), consisting of three color components of image pixels in the Lab color space and the two-dimensional position coordinates of each pixel in the Cartesian coordinate system. Then, the distance metric of the five-dimensional vector is constructed to cluster superpixels iteratively. To achieve better separation performance and balance the running time of the algorithm, it is also necessary to determine the number of split superpixels K and compactness L.

The optimal number of split superpixels K is generally determined based on the actual separation accuracy. In this study, the value of K also needs to be determined in conjunction with whether the superpixel skewness information couldcan be effectively expressed. The undersegmentation error (*Ue*) was selected as the evaluation index for separation accuracy. *Ue* describes the proportion of pixels that exceed the true data boundary [29], and the specific formula is as follows:

$$Ue = \frac{\sum_i \left(\sum_{S_j | S_j \cap G_i > B} |S_j| \right) - N}{N} \tag{1}$$

,where $S_1 \ldots S_n$ indicates N superpixel blocks generated by the superpixel algorithm; G is the reference area; B controls the percentage threshold of superpixels generated by the superpixel segmentation algorithm that overlaps with the reference segmentation area, with a value of 5%; i is the sequence number of the benchmark segmentation area; and j is the serial number of the superpixel region.

The tightness L of superpixel clustering, which ranges from 0 to 40, is another important parameter of SLIC. The smaller the tightness value, the finer the boundary between superpixels. In this study, the value is 1.

Extraction of skewness parameters in narrowband spectral images

After discovering the skewed distribution of superpixel color gradation in narrowband spectral images, this study introduces the RGB color model skewed distribution mode to expand the superpixel feature parameters of narrowband spectral images. This study specifically referred to the method of Chen et al. [24] and used the corresponding Python library functions to extract skewed distribution parameters from narrowband spectral canopy images. A total of 20 skewed leaf color parameters were obtained, including the mean (R_{Mean}), median (R_{Median}), mode (R_{Mode}), skewness ($R_{Skewness}$), and kurtosis (R_{Kurtosis}) of the R-channel color scale, the mean (G_{Mean}), median (G_{Median}), mode (G_{Mode}), skewness ($G_{Skewness}$), and kurtosis (G_{Kurtosis}) of the G-channel color scale, and the mean (B_{Mean}), median (B_{Median}), mode (B_{Mode}), skewness ($B_{Skewness}$), and kurtosis (B_{Kurtosis}) of the B-channel color scale, as well as the mean (Y_{Mean}), median (Y_{Median}), mode (Y_{Mode}), skewness ($Y_{Skewness}$), and kurtosis (Y_{Kurtosis}) of the Y-image color scale.

This study used the corresponding Python library functions to extract Lab color model parameters and HSV color model parameters from narrowband spectral images, resulting in a total of 26 superpixel feature parameters in three categories (Table 1). In the process of extracting superpixel feature parameters, Python libraries such as cv2, dnumpy, Scipy, and skimage.feature were

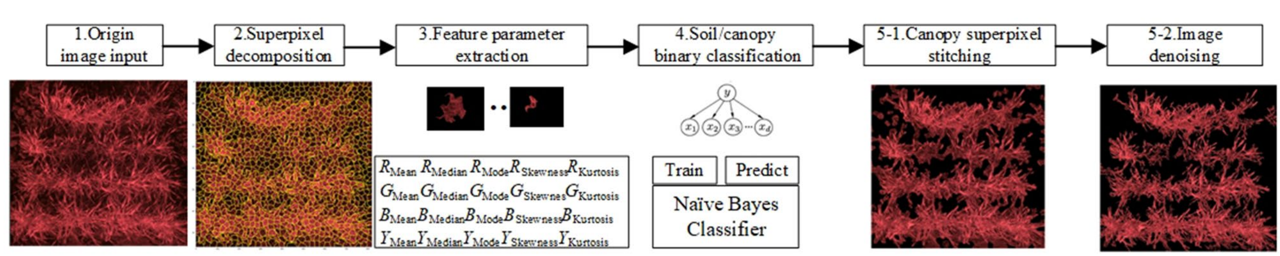

Fig. 2 Narrowband spectral image canopy extraction process

Table 1 Superpixel feature parameters of narrowband spectral images

Parameter system	Parameter indicators	Number of parameters
Color model skewness parameters of RGB	R_{Mean}, R_{Median}, R_{Mode}, $R_{Skewness}$, $R_{Kurtosis}$; G_{Mean}, G_{Median}, G_{Mode}, $G_{Skewness}$, $G_{Kurtosis}$; B_{Mean}, B_{Median}, B_{Mode}, $B_{Skewness}$, $B_{Kurtosis}$; Y_{Mean}, Y_{Median}, Y_{Mode}, $Y_{Skewness}$, $Y_{Kurtosis}$	20
Color model skewness parameters of Lab	*L*, *a*, *b*	3
Color model skewness parameters of HSV	*H*, *S*, *V*	3

mainly used in the process of extracting superpixel image parameters.

Construction of bayesian classifier based on skewed feature parameters

The classification principle of the Bayesian classifier is to calculate the posterior probability of an object using a Bayesian formula based on its prior probability; that is, the probability that the object belongs to a certain class. The class with the highest posterior probability is selected as the class to which the object belongs. In other words, Bayesian classifiers are optimized in terms of the minimum error rate. Currently, four types of Bayesian classifiers have been studied extensively: Naive Bayes, TAN, BAN, and GBN. In this study, the Naive Bayes type was selected [30, 31]. The advantage of Bayesian classifiers is that the prediction process is simple and fast, which makes them suitable for the development of embedded systems. They are also effective for multiclassification problems, and the complexity does not increase significantly.

Construction of bayesian classification model To ensure that the constructed Bayesian classifier had good light intensity adaptability, six images with different illumination conditions from the 48 collected images were selected as the training set, and the remaining 42 images were assigned into the test set. The training set images were split according to the optimal K value obtained as described in Sect. 2.2.2 of this study, and 100 typical canopy layers and 100 typical soil superpixels were selected from each image. A total of 600 canopy layers and 600 soil narrowband spectral image superpixels were selected for the entire training set, and 26 feature parameters were calculated for each superpixel in the training set (as shown in Table 1) as the input for the training model. The classifier used the canopy and background as two-class outputs, and was trained using the Python programming language and an NVIDIA Jetson TX2 processor to quickly construct a Bayesian classifier model quickly.

Canopy superpixel stitching and denoising

The original image to be tested is subjected to superpixel decomposition (the best K value obtained in Sect. 2.2.2), and 26 feature parameters are extracted sequentially according to the index position of the superpixels in the original image. Then, they are input into the trained Bayesian classification model. If they are judged as a canopy, they enter the superpixel stitching step. This cycle continues until all superpixels in an image are predicted.

After the narrowband spectral image superpixels are identified using a Bayesian classifier, all canopy superpixels can be concatenated to generate canopy extraction images with rough boundaries. The boundary of the image extracted from the canopy may contain a misclassified soil background. Therefore, it is necessary to denoise the concatenated image. The denoising algorithm for the canopy mosaic images uses the rgb2gray function to convert the mosaic images into grayscale images. The Otsu method is used to determine the thresholds and binarize the grayscale images. In addition, the label function is used to label all connected regions and the regionprops method is used to calculate the size of each region. After removing connected regions of less than 2200 pixels, a masked image without small areas is obtained, and the original image is intersected with the mask to obtain an optimized canopy image that removes noise interference.

Evaluation of canopy extraction performance in narrowband spectral images based on superpixel color gradation skewness distribution features

Test set source

The remaining 42 narrowband spectral images were obtained as described in Sect. 2.2.4 as the test set, and four narrowband spectral images under different illumination conditions were selected for manual and accurate labeling (as shown in Fig. 3). This study used the manually labeled results as a reference to facilitate the display and evaluation of the effectiveness of canopy extraction. The relevant evaluation indicators are described in Sect. 2.3.2. In addition, the test set of classifiers with different parameters in Sect. 2.4 is based on these four images.

Evaluation indicators for canopy extraction

The image segmentation effect can be evaluated by the intersection over union (IoU) [32], and the calculation method is shown in formula (2). The IoU value reflects the degree of consistency between the segmentation result and true value, with a value between 0 and 1. The higher the IoU value, the better the segmentation effect.

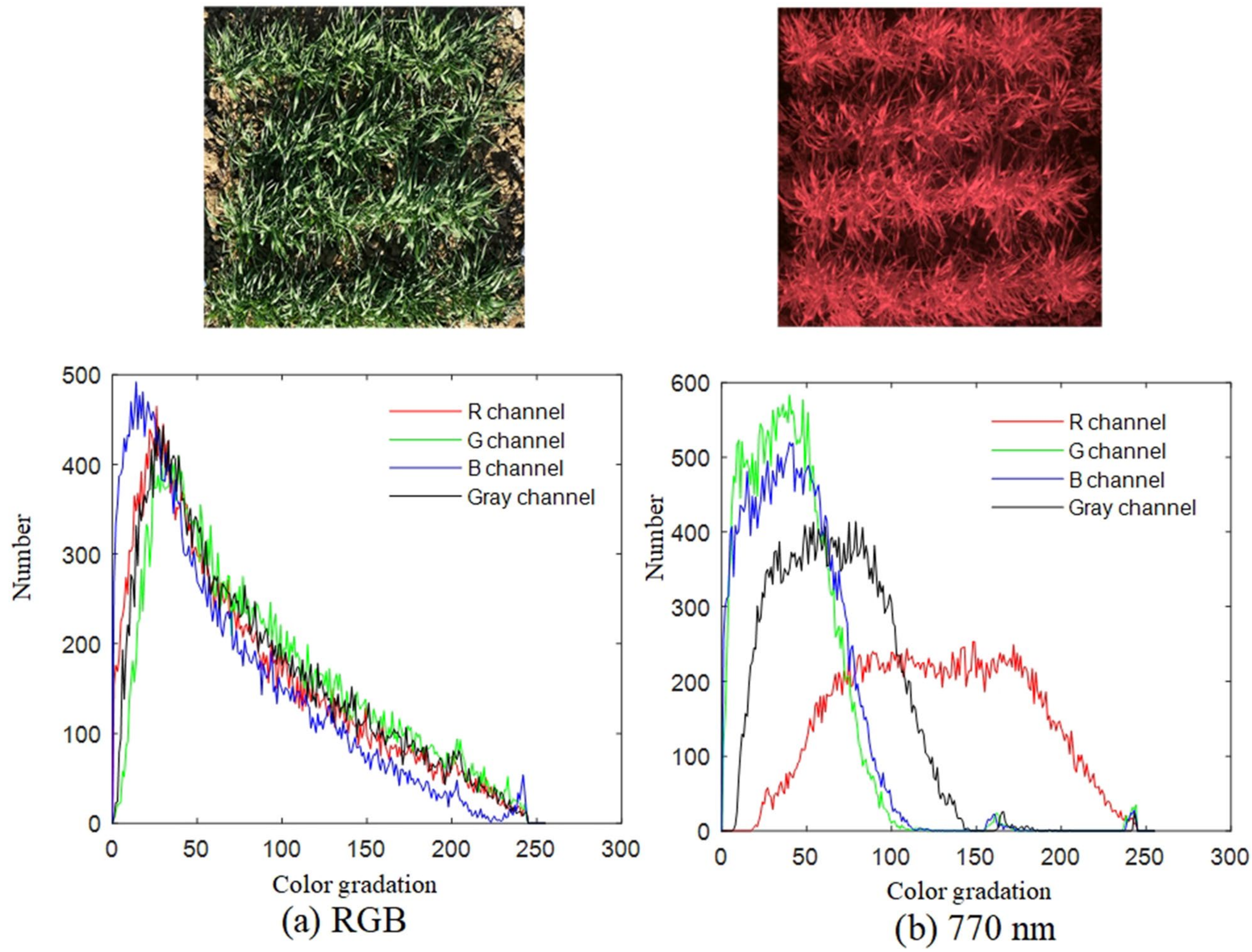

Fig. 3 Canopy extraction effect of narrowband spectral images based on skewed distribution parameters

$$IoU = \left(\frac{\sum_{i=0}^{n}\sum_{j=0}^{m}\left(A(v)_{i,j} \cap B(v)_{i,j}\right)}{\sum_{i=0}^{n}\sum_{j=0}^{m}\left(A(v)_{i,j} \cup B(v)_{i,j}\right)} \right) *100\% \tag{2}$$

,where $A(v)$ represents the set of pixel categories segmented by the segmentation algorithm, including the background $(v=0)$ and foreground $(v=1)$; $B(v)$ is the true label set corresponding to the pixel set, including the background $(v=0)$ and foreground $(v=1)$; i, j are the pixel index; *m* is the height of the image; *n* is the width of the image; and $v_{i,j}$ is the grayscale value of the pixel in the *i*-th column and *j*-th row.

Comparison of classification accuracy of narrowband spectral superpixel classifiers with different input parameters

Construction of narrowband spectral superpixel classifiers with different input parameters

To compare the Bayesian classification accuracy under different parameter systems and further clarify the role and significance of skewed distribution parameters in this study, the four images obtained in Sect. 2.3.1, were considered as processing objects, Bayesian classifier models Y1 and Y2 were constructed by inputting different types of parameter combinations, and the classification accuracy of Y1 and Y2 were evaluated using the same training and testing sets. The relevant parameters of the Y1 and Y2 models are shown in Table 2, where Train-Data is the training set described inSection 2.2.4 (600 canopy layers and 600 soil narrowband spectral image superpixels); Test Data is the set of these four images under the optimal number of K (each image was split into 1000 superpixels, with a total of 4000 superpixels for the four images).

Evaluation indicators for model classification accuracy

The classification accuracy index Acc of the Bayesian classifier is defined in formula (3).

$$Acc = \frac{TP + TN}{TN + FN + TP + FP} \tag{3}$$

Table 2 Bayesian classifier modeling with different input parameters

Mode	Train_Data	Test_Data	Input parameters	Number of input parameters
Y1	1200	4000	R_{Mean}; R_{Median}; R_{Mode}; G_{Mean}; G_{Median}; G_{Mode} B_{Mean}; B_{Median}; B_{Mode};; Y_{Mean}; Y_{Median}; Y_{Mode} L; a; b; H; S; V	18
Y2	1200	4000	R_{Mean}; R_{Median}; R_{Mode}; $R_{Skewness}$; $R_{Kurtosis}$ G_{Mean}; G_{Median}; G_{Mode}; $G_{Skewness}$; $G_{Kurtosis}$ B_{Mean}; B_{Median}; B_{Mode}; $B_{Skewness}$; $B_{Kurtosis}$ Y_{Mean}; Y_{Median}; Y_{Mode}; $Y_{Skewness}$; $Y_{Kurtosis}$ L; a; b; H; S; V	26

,where TP is the number of superpixels with the true category as the canopy and the predicted category as the canopy;

FN indicates that the true category of superpixels is the canopy and the predicted category is the number of soil backgrounds;

FP indicates that the true category of the superpixels is the soil background, and the predicted category is the number of canopy layers;

TN indicates that the true category of superpixels is the soil background and the predicted category is the number of soil backgrounds.

Results and analysis

Narrowband spectral image canopy leaf color skewness distribution

Leaf color skewness phenomenon in narrowband spectral images at canopy scale

This study analyzed the leaf color gradation of closed-row wheat canopy images collected using RGB color cameras and 770 nm narrowband spectral cameras. A skewed distribution of leaf color gradation was observed in the narrowband spectral images (Fig. 4). We conducted a skewed distribution check on the Lilliefors test of the R, G, B, and Gray channels of the two images, as shown in Fig. 4. This study found that the distribution of leaf color gradation in the four channels of the two images conformed to a skewed distribution and was significant, with an H value of 1 and $P<0.01$, which rejects the hypothesis of a normal distribution.

Skewness test based on superpixel narrowband spectral images

In Sect. 3.1.1, it was found that a skewed distribution of leaf color gradation exists in narrowband spectral images. Therefore, this study aimed to further determine whether each superpixel in narrowband spectral images also exhibits a skewed distribution of leaf color gradation at the superpixel scale. Six 770 nm narrowband spectral images with a size of 3072×2048 pixels were selected and each image was split into 200, 400, 600, 800, 1000, and 1200 superpixels. Figure 5 shows the splitting effects of the six superpixels of different sizes representing the images.

Among the six images, 4200 superpixels of different sizes were subjected to the Lilliefors test for the R, G, B, and Gray channels individually. Out of the 4200 superpixels, 4190 conformed to the leaf color gradient and skewness distribution, which was significant, with an H value of 1 and $P<0.01$, thereby negating the hypothesis of a normal distribution. The statistical results are presented in Table 3.

After individually examining the remaining 10 superpixels that did not conform to a skewed distribution, the study found that there were two main reasons why superpixels exhibited non-skewed distributions. The first type of superpixel clustering failure occurred when the superpixels clustered the soil and canopy together, as shown in Fig. 6(a). The color distribution of the soil background and canopy presented a bimodal structure, and there was no boundary between them, making it impossible to set a threshold for segmentation. This situation was extremely rare, with only one case in 4200 superpixels. The second type, as shown in Fig. 6(b), was owing to the improper selection of the segmentation threshold for the soil background (the threshold was uniformly set to be greater than 35), resulting in incomplete removal of the color scale of the background interference and failure of the skewed distribution test. After resetting the threshold, all nine superpixels in Fig. 6(b) conformed to the skewed distribution. Therefore, preprocessing of superpixels (removing background interference) before skewed distribution verification is crucial. Overall, the criterion for the narrowband spectral image leaf color to conform to a skewed distribution is robust.

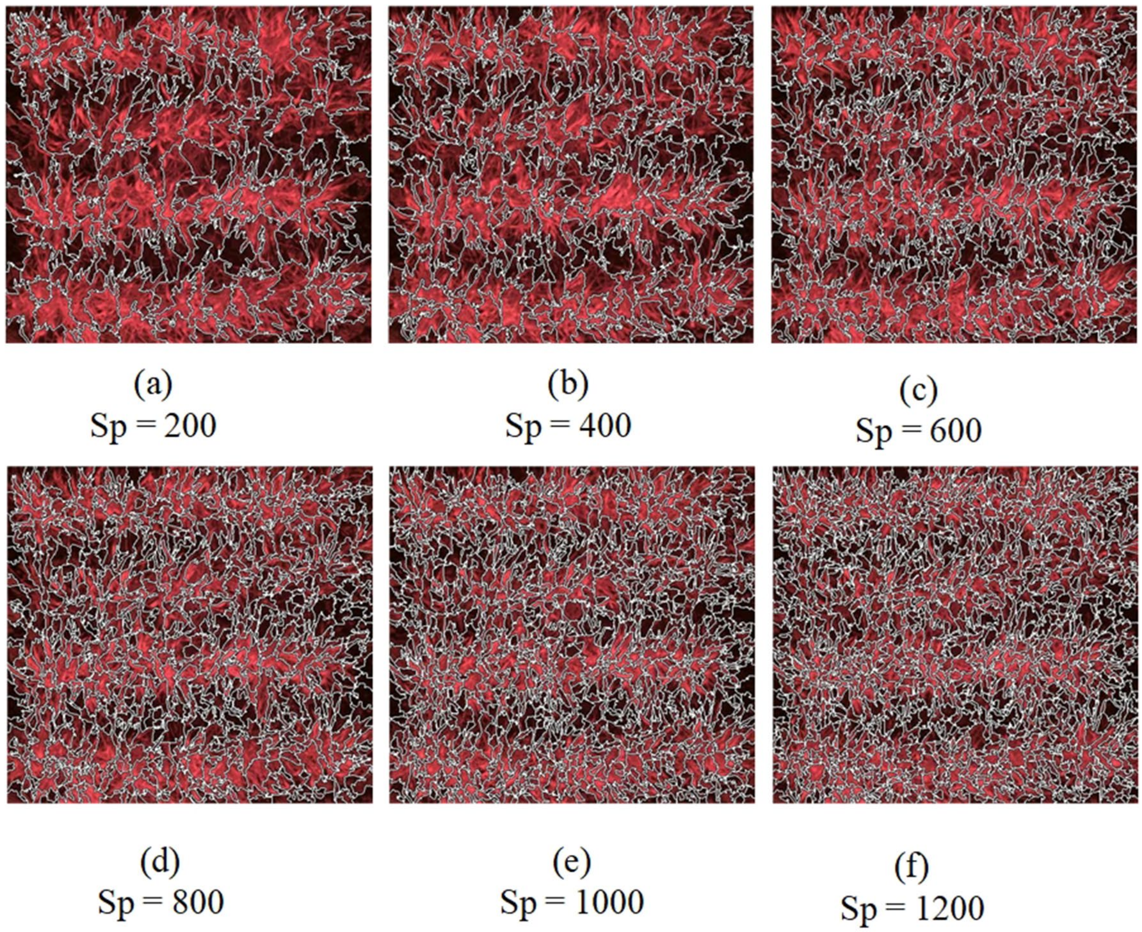

Fig. 4 The skewed distribution phenomenon of leaf color in narrowband spectral images

Determination of superpixel parameters

The determination *K* of the number of superpixel segments in the SLIC algorithm requires a comprehensive consideration of two factors. First, the *K* value needs to meet the segmentation accuracy requirements of *Ue* described in Sect. 2.2.2; second, it is also necessary to consider whether a large *K* value can effectively express skewed information, as a single superpixel contains less pixel information.

To express the skewed information of superpixels of different sizes, Fig. 7 shows the distribution of the leaf color gradient of superpixel R channels with different numbers of cracks in each image. The leaf color gradient showed a skewed distribution, and as the number of cracks increased, the number of burrs in the color scale distribution increased. This is mainly owing to the increase in the number of cracks, which reduces the pixel information contained in a single superpixel and increases the fluctuation of the color scale. Therefore, when performing superpixel decomposition, attention should be paid to the superpixel size to avoid losing skewed information if it is too small. According to the data performance shown in Fig. 7, the *K* value of the number of superpixel segments in the narrowband spectral image canopy extraction algorithm should not exceed 1200.

To meet the segmentation accuracy of *Ue*, four 770 nm narrowband spectral images with a resolution of 3072×2048 were selected and accurately labeled. The selected images were divided into 200, 400, 600, 800, 1000, and 1200 sequences and the segmentation accuracy of *Ue* is shown in Fig. 8. The results indicate that when *K* was 1000, the average value of *Ue* was at a local minimum; if the *K* value was too large, the segmentation of superpixels was prone to losing its meaning. Therefore, considering these two factors, the *K* value of the number of superpixel segments was set to 1000.

Segmentation performance of narrowband spectral image canopy extraction method based on superpixel color gradient skewed distribution characteristics

In Fig. 3, the number of superpixel splits *K* for the four images was 1000, the tightness was 1, and the Bayesian classification model was Y2. Following the narrowband spectral image canopy extraction process shown in Fig. 2,

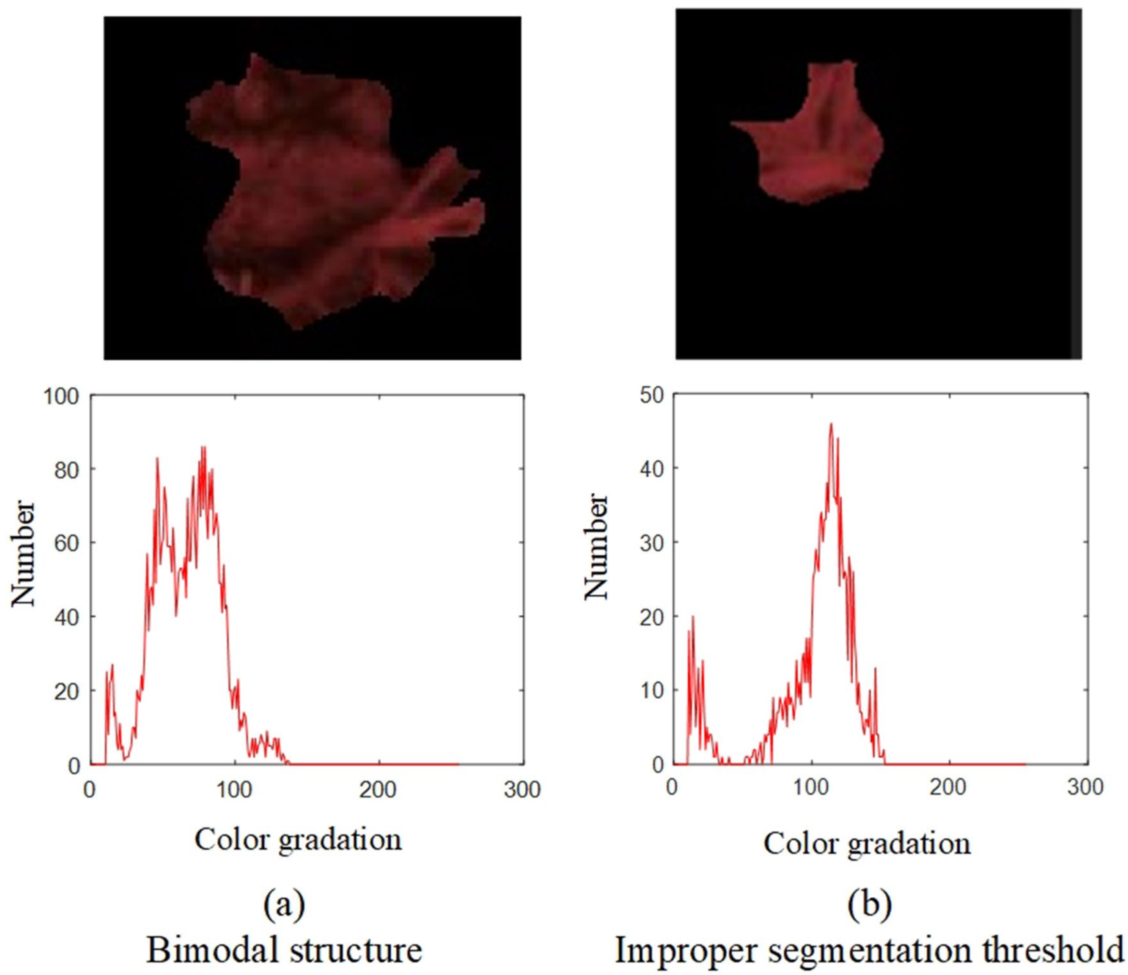

Fig. 5 Narrowband spectral image splitting by different superpixel sizes

the average IoU for the four images was 90.41% and the average IoU for Otsu was 77.97%. The actual canopy extraction performance of the four images was significantly better than that of the traditional Otsu processing methods, and images captured in different light environments had good adaptability.

Performance comparison of narrowband spectral superpixel classifiers with different input parameters

As shown in Fig. 3, the four images generated 4000 superpixels. The actual classification accuracies based on the Y1 and Y2 Bayesian classifiers are listed in Table 4. Acc_Y1 (the average classification accuracy of the Y1 model) was 72.02% and Acc_Y2 (the average classification accuracy of the Y2 model) was 91.76%. The performance results of the Y1 and Y2 classifiers indicate that the leaf color gradient skewness parameter introduced in this study can significantly improve the classification accuracy of narrowband spectral images of the canopy and soil.

Discussion

This study investigated the skewed distribution of leaf color gradation in narrowband spectral images, introduced RGB color space skewed distribution parameters, and proposed a narrowband spectral image canopy extraction method based on superpixel color gradation skewed distribution characteristics. The main contributions of this study are as follows:

(1) This study verified that narrowband spectral images and their superpixel color gradations conform to skewed distributions.

After testing the color distribution of narrowband wheat spectral canopy images, we found that narrowband spectral images and their split superpixel color distributions followed a skewed distribution. This characteristic was used to expand the superpixel feature parameters of the narrowband spectral images and introduce RGB color space skewed distribution parameters, as shown in Table 1. Because superpixels are commonly used processing objects in feature parameter decision-making methods, this study further explored whether superpixels

Table 3 Skewness test for different superpixel sizes in four channels

Normal distribution test	Sample size	*R* channel			G channel			B channel			Gray channel		
		H	*P* < 0.05	*P* < 0.01	H	*P* < 0.05	*P* < 0.01	H	*P* < 0.05	*P* < 0.01	H	*P* < 0.05	*P* < 0.01
Lilliefors test	200	1	100%	100%	1	100%	100%	1	100%	100%	1	100%	100%
	400	1	100%	100%	1	100%	100%	1	100%	100%	1	100%	100%
	600	1	100%	100%	1	100%	100%	1	100%	100%	1	100%	100%
	800	1	99.75%	99.75%	1	99.75%	99.75%	1	99.75%	99.75%	1	99.75%	99.75%
	1000	1	99.70%	99.70%	1	99.98%	99.97%	1	99.98%	99.97%	1	99.98%	99.97%
	1200	1	99.67%	99.58%	1	99.67%	99.58%	1	99.67%	99.58%	1	99.67%	99.58%

of different sizes also conformed to a skewed distribution, which is of great value for the practical application of narrowband spectral leaf color skewed distribution characteristics. The narrowband spectral images were split according to *K* values of 200, 400, 600, 800, 1000, and 1200, and the Lilliefors test skewness distribution verification was performed on the selected 4200 split superpixels individually. It was found that 4190 superpixels had leaf colors that matched the skewness distribution characteristics, whereas the remaining superpixels that did not match the skewness distribution characteristics were mainly caused by soil background interference. Therefore, the preprocessing of superpixels before skewness distribution verification (removing background interference) is also crucial. Overall, the criterion for the skewed distribution of leaf color in narrowband spectral images is robust.

(2) This study clarified the precautions for the application of superpixel skewed distribution.

It was also noted that, as the number of cracks increased, the number of burrs in the color gradient distribution increased. This may be owing to the increase in the number of cracks, which reduced the pixel information contained in a single superpixel and increased the fluctuation of the color scale. This phenomenon inspired us to pay attention to the superpixel size when performing superpixel decomposition to avoid losing skewed information when the superpixel size was too small. In the actual selection of superpixel *K* values, consideration should be given to the accuracy requirements of the superpixel segmentation *Ue* as well as the expression of superpixel skewness information.

Based on the above two innovative points, this study explored the application rules of leaf color skewness distribution characteristics in narrowband spectral images, providing a theoretical basis for narrowband spectral image canopy extraction methods based on superpixel size color skewness distribution characteristics. In the actual narrowband spectral image canopy extraction process, the method based on superpixel size gradient skewness distribution features achieved a segmentation accuracy of 90.41%, which was significantly higher than that of the traditional Otsu processing method (77.97%), and this method had good adaptability to images captured in different light environments. At the application level of practical feature parameters, introducing leaf color gradient skewness parameters (skewness and kurtosis) can significantly improve the classification accuracy of narrowband spectral images of the canopy and soil.

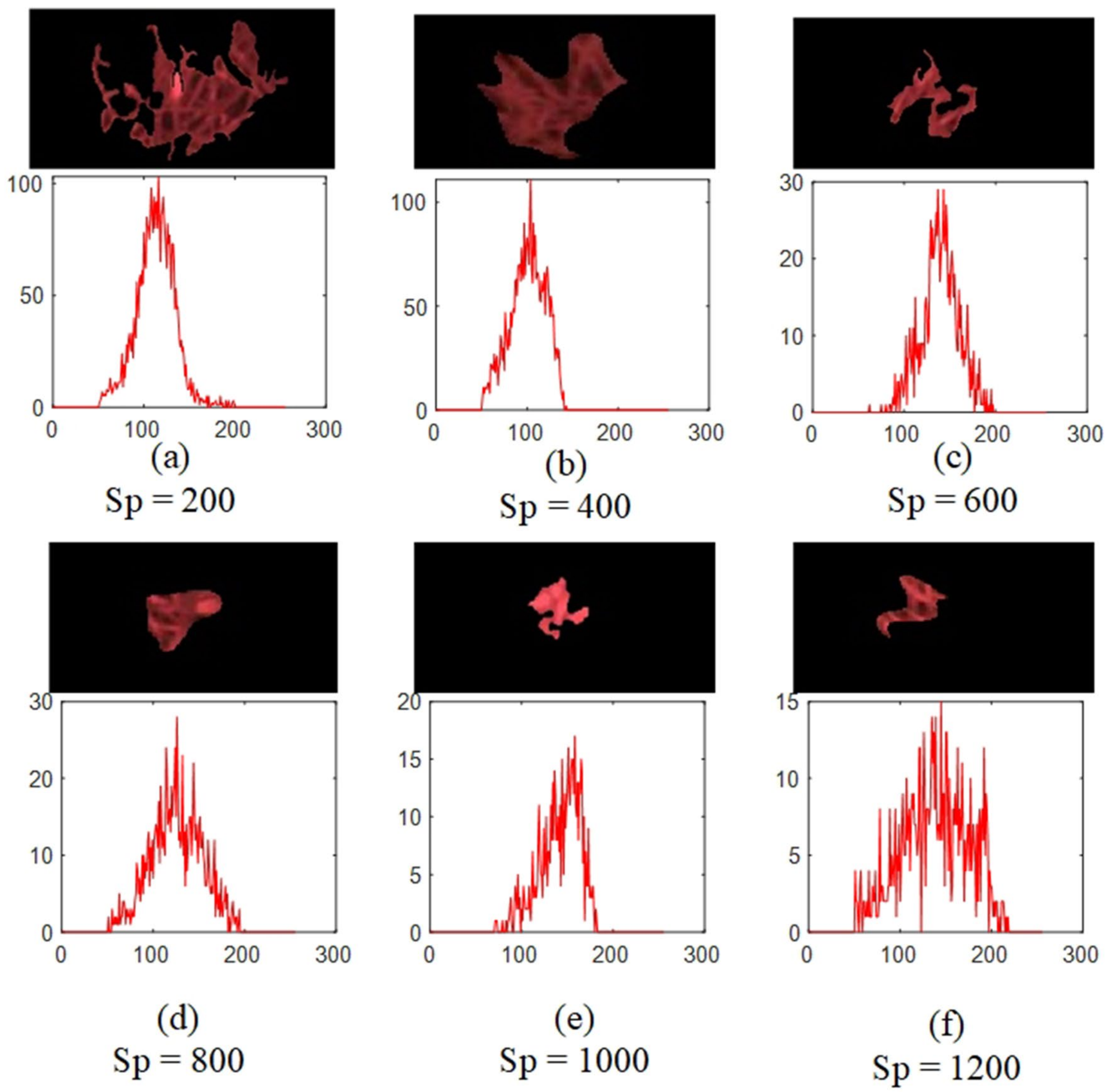

Fig. 6 Non-skewed distribution analysis of superpixels

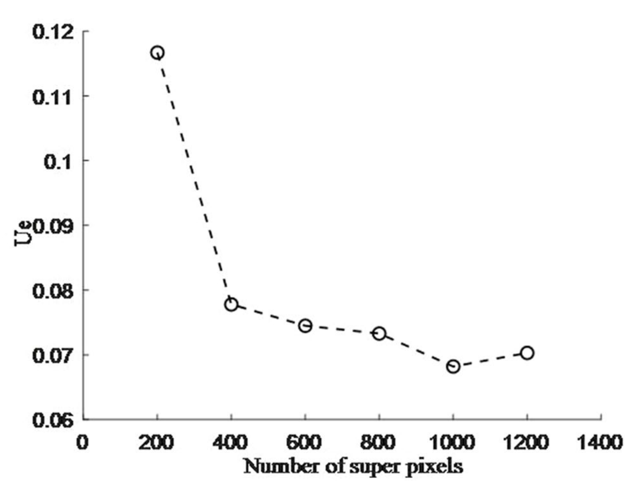

Fig. 7 Display of skewed distribution of different superpixel sizes

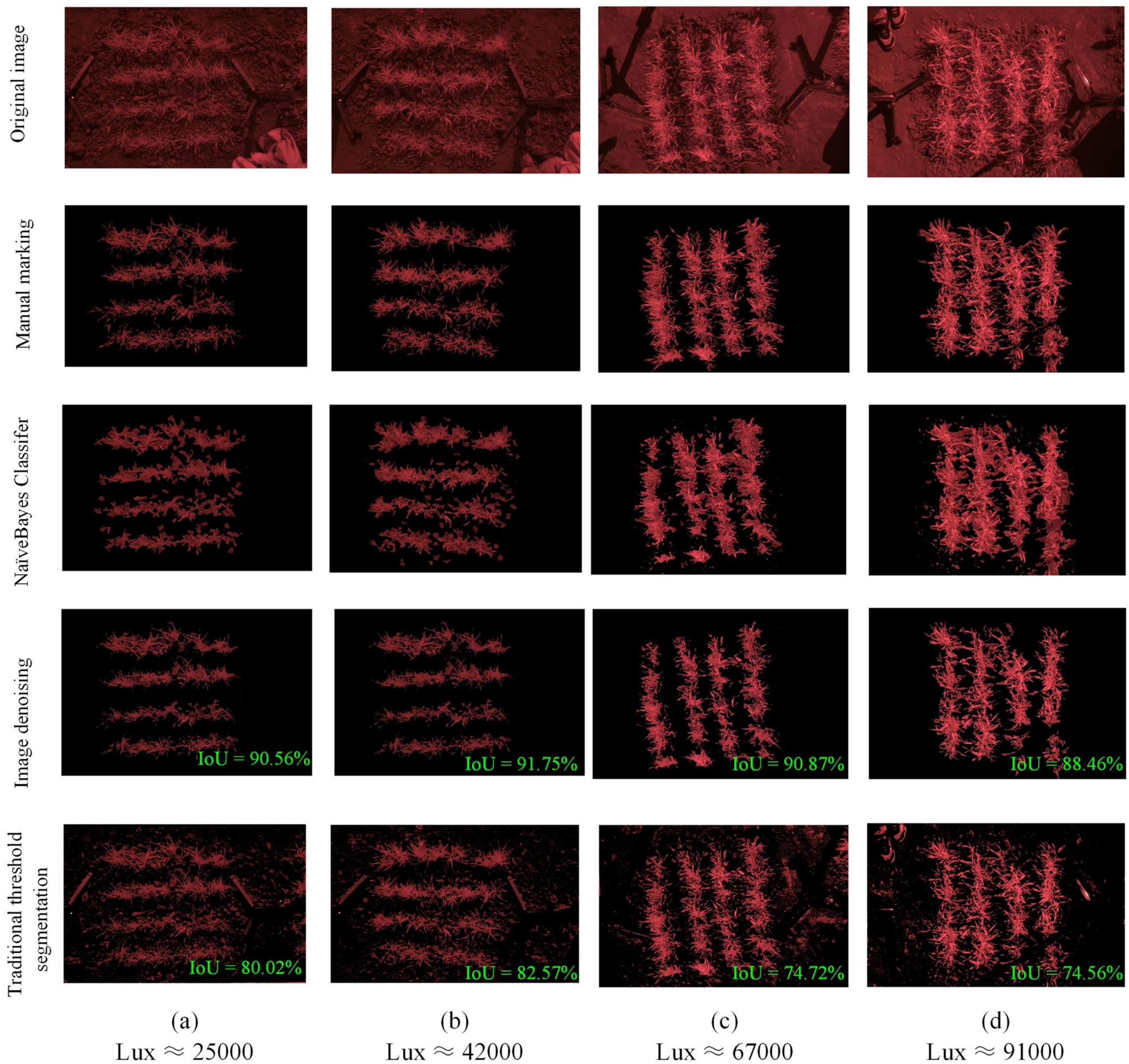

Fig. 8 Segmentation accuracy of different numbers of superpixels *K*

Table 4 Accuracy of superpixels

Objects	Acc_Y1	Acc_Y2
a	69.4%	89.7%
b	71.3%	93.2%
c	73.4%	90.3%
d	73.8%	93.6%

Conclusions

Narrowband spectral images and their decomposed superpixel leaf color scales follow a skewed distribution that expands the expression of leaf color information in narrowband spectral images. Compared with traditional threshold segmentation methods, the narrowband spectral image canopy extraction method based on superpixel size gradient skewness distribution features has higher segmentation performance and light intensity adaptability. The leaf color skewness feature parameters in this study are empirical and robust and can be used to construct a classifier model with relatively simple hardware and fewer samples, which will greatly reduce the modeling process and make it more suitable for embedded development in real-time field collection scenarios.

Author contributions

Conceptualization, YD and HY; methodology, PZ and HY; software, HY and ZC; formal analysis, HY and RZ; data curation, HY and XD; writing—original draft preparation, YD and HY. writing— review and editing, PZ and YD. All authors read and approved the final manuscript.

Funding
This work was mainly supported by the Jiangsu Province Key Technologies R&D Program (BE2023368), Shandong Province Key Technologies R&D Program (2022SFGC0203), and Natural Science Foundation of Jiangsu Province (BK20241581).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors agreed to publish this manuscript.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Engineering, Nanjing Agricultural University, Nanjing 210031, China
[2]College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China
[3]Collaborative Innovation Center for Modern Crop Production Co-sponsored by Province and Ministry, Nanjing Agricultural University, Nanjing 210095, China
[4]Jiangsu Meteorological Bureau, Nanjing 210008, China
[5]Longyan Company of Fujian Provincial Tobacco Corporation, Longyan 364000, People's Republic of China

Received: 1 July 2024 / Accepted: 23 September 2024
Published online: 01 October 2024

References
1. Feng A, Zhou J, Vories E, Sudduth KA. Evaluation of cotton emergence using UAV-based narrow-band spectral imagery with customized image alignment and stitching algorithms. Remote Sens. 2020;12:1764. https://doi.org/10.3390/rs12111764.
2. Weng H, Tian Y, Wu N, Li X, Yang B, Huang Y, et al. Development of a low-cost narrow band multispectral imaging system coupled with chemometric analysis for rapid detection of rice false smut in rice seed. Sens (Basel). 2020;20:1209. https://doi.org/10.3390/s20041209.
3. Li Y, Chen J, Nie J, Li J, Ercisli S. Low-carbon jujube moisture content detection based on spectral selection and reconstruction. IEEE Internet Things J. 2024;1. https://doi.org/10.1109/JIOT.2024.3368140.
4. Zhang J, Wang C, Yang C, Jiang Z, Zhou G, Wang B, et al. Evaluation of a UAV-mounted consumer grade camera with different spectral modifications and two handheld spectral sensors for rapeseed growth monitoring: performance and influencing factors. Precision Agric. 2020;21:1092–120. https://doi.org/10.1007/s11119-020-09710-w.
5. Yang J, Guo X, Li Y, Marinello F, Ercisli S, Zhang Z. A survey of few-shot learning in smart agriculture: developments, applications, and challenges. Plant Methods. 2022;18:28. https://doi.org/10.1186/s13007-022-00866-2.
6. Zhou J, Wang B, Fan J, Ma Y, Wang Y, Zhang Z. A systematic study of estimating potato N concentrations using UAV-based hyper- and multi-spectral imagery. Agronomy. 2022;12:2533. https://doi.org/10.3390/agronomy12102533.
7. Yang X, Wang X, Evers B, Cui Y, Poland J. Spectral correlation between wheat genotype replications over the visible and near-infrared spectrum. Remote Sens Lett. 2022;13:1203–12. https://doi.org/10.1080/2150704X.2022.2138619.
8. Fang Y, Wang P, Chen J, Tian Q. Simulation and analysis on the influence of different types of soil background on the remote sensing information of wheat NDVI of farmland. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2016. IEEE Publications; 2016 p. 4462–5. https://doi.org/10.1109/IGARSS.2016.7730163
9. Prudnikova E, Savin I, Vindeker G, Grubina P, Shishkonakova E, Sharychev D. Influence of soil background on spectral reflectance of winter wheat crop canopy. Remote Sens. 2019;11:1932. https://doi.org/10.3390/rs11161932.
10. Xing Z. An improved emperor penguin optimization based multilevel thresholding for color image segmentation. Knowl Based Syst. 2020;194:105570. https://doi.org/10.1016/j.knosys.2020.105570.
11. Wang Y, Wang D, Zhang G, Wang J. Estimating nitrogen status of rice using the image segmentation of G-R thresholding method. Field Crops Res. 2013;149:33–9. https://doi.org/10.1016/j.fcr.2013.04.007.
12. Yu HF, Xu HL, Ding YQ, Yang ZN, Dou XL, Li Q et al. Wheat canopy vegetation index measurement method based on narrow band spectral image analysis; 2023.
13. Nie J, Wang Y, Li Y, Chao X. Artificial intelligence and digital twins in sustainable agriculture and forestry: a survey. Turk J Agric for. 2022;46:642–61. https://doi.org/10.55730/1300-011X.3033.
14. Nie J, Wang Y, Li Y, Chao X. Sustainable computing in smart agriculture: survey and challenges. Turk J Agric for. 2022;46:550–66. https://doi.org/10.55730/1300-011X.3025.
15. Rose S, Nickolas S, Sangeetha S. Machine learning and statistical approaches used in estimating parameters that affect the soil fertility status: A survey. In: 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE Publications; 2018 p. 381–5. https://doi.org/10.1109/ICGCIoT.2018.8753025
16. Dan S. NIR spectroscopy oranges origin identification framework based on machine learning. Int J Semant Web Inf Syst. 2022;18:1–16. https://doi.org/10.4018/IJSWIS.297039.
17. Lu Y, Yang H, Sun A. The research of Spad in rice leaves based on machine learning. In: 2019 Chinese Automation Congress (CAC). IEEE Publications; 2019 p. 2163–7. https://doi.org/10.1109/CAC48633.2019.8996863
18. Zhang J, Wang P, Gong F, Zhu H, Chen N. Content-based superpixel segmentation and matching using its region feature descriptors. IEICE Trans Inf Syst. 2020;E103D:1888–900. https://doi.org/10.1587/transinf.2019EDP7322.
19. Singh AK, Sreenivasu SVN, Mahalaxmi USBK, Sharma H, Patil DD, Asenso E. Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier. J Food Qual. 2022;2022:1–16. https://doi.org/10.1155/2022/2845320.
20. Lu H, Cao ZG, Xiao Y, Fang Z, Zhu Y, Xian K. Fine-grained maize tassel trait characterization with multi-view representations. Comput Electron Agric. 2015;118:143–58. https://doi.org/10.1016/j.compag.2015.08.027.
21. Yang Q, Chen Y, Xun Y, Bao G. Superpixel-based segmentation algorithm for mature citrus. Int J Agric Biol Eng. 2020;13:166–71. https://doi.org/10.25165/j.ijabe.20201304.5607.
22. Xiong X, Duan LF, Liu LB, Tu H, Yang P, Wu D, et al. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. Plant Methods. 2017;13:104. https://doi.org/10.1186/s13007-017-0254-7.
23. Li Y, Chao X. Distance-entropy: an effective indicator for selecting informative data. Front Plant Sci. 2021;12:818895. https://doi.org/10.3389/fpls.2021.818895.
24. Chen ZM, Wang FZ, Zhang P, Ke C, Zhu Y, Cao W, et al. Skewed distribution of leaf color RGB model and application of skewed parameters in leaf color description model. Plant Methods. 2020;16:23. https://doi.org/10.1186/s13007-020-0561-2.
25. Wu X, Zhang F, Lv J. Research on recognition of tea tender leaf based on image color information. J Tea Sci. 2013;33:584–9.
26. Han WT, Sun Y, Xu TF, Chen XW, Su KO. Detecting maize leaf water status by using digital RGB images. Int J Agric Biol Eng. 2014;7:45–53.
27. Choi KS, Oh KW. Fast simple linear iterative clustering for superpixel segmentation. In: 2015 IEEE International Conference on Consumer Electronics (ICCE). IEEE Publications; 2015 p. 552–3.
28. Al-Azawi RJ, Al-Jubouri QS, Mohammed YA. Enhanced algorithm of superpixel segmentation using simple linear iterative clustering. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE). IEEE Publications; 2019 p. 160–3. https://doi.org/10.1109/DeSE.2019.00038
29. Zhao YQ, Chen GS, Jing WP. Improved SLIC clustering superpixels segmentation method on remote sensing image. J Northeast Univ. 2020;48:66–71. (in Chinese with English abstract).
30. Winarti T, Vydia V. Feature selection for optimizing the Naive Bayes algorithm. In engineering, information and agricultural technology in the global digital revolution. In: Proceedings of the 1st international conference on civil engineering, electrical engineering, information systems, information technology,

and agricultural technology (SCIS 2019), Jul 10, 2019, Semarang, Indonesia. CRC Press; 2020 (p. 47).
31. Chen S, Webb GI, Liu L, Ma X. A novel selective naïve Bayes algorithm. Knowl Based Syst. 2020;192:105361. https://doi.org/10.1016/j.knosys.2019.105361.
32. Duan LF, Xiong X, Liu Q, Yang WN, Huang CL. Field rice panicles segmentation based on deep full convolutional neural network. Trans Chin Soc Agric Eng (Transactions of the CSAE). 2018;34:202–9 (in Chinese with English abstract). Merops.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.