12-11-2021

GIS-based volunteer cotton habitat prediction and plant-level detection with UAV remote sensing

Tianyi Wang

Xiaohan Mei

J. Alex Thomasson

Chenghai Yang

Xiongzhe Han

See next page for additional authors

Authors

Tianyi Wang, Xiaohan Mei, J. Alex Thomasson, Chenghai Yang, Xiongzhe Han, Pappu Kumar Yadav, and Yeyin Shi

Computers and Electronics in Agriculture 193 (2022) 106629

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

GIS-based volunteer cotton habitat prediction and plant-level detection with UAV remote sensing

Tianyi Wang [a,b,*], Xiaohan Mei [c], J. Alex Thomasson [b,d], Chenghai Yang [e], Xiongzhe Han [f,g,*], Pappu Kumar Yadav [b], Yeyin Shi [h]

[a] College of Engineering, China Agricultural University, P.O. Box 134, No. 17 Qinghua East Road, Haidian District, Beijing 100083, China
[b] Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843, USA
[c] Department of Water Management and Hydrological Science, Texas A&M University, College Station, TX 77843, USA
[d] Department of Agricultural and Biological Engineering, Mississippi State University, Mississippi State, MS 39762, USA
[e] USDA-Agricultural Research Service, Aerial Application Technology Research Unit, College Station, TX 77845, USA
[f] Department of Biosystems Engineering, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon 24341, Republic of Korea
[g] Interdisciplinary Program in Smart Agriculture, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon 24341, Republic of Korea
[h] Department of Biological Systems Engineering, University of Nebraska, Lincoln, NE 68583, USA

ARTICLE INFO

Keywords:
Volunteer cotton
Feral cotton
GIS
Network analysis
Plant-level
Classification
Precision agriculture
Cotton boll weevil

ABSTRACT

Volunteer cotton plants germinate and grow at unwanted locations like transport routes and can serve as hosts for a harmful cotton pests called cotton boll weevils. The main objective of this study was to develop a geographic information system (GIS) framework to efficiently locate volunteer cotton plants in the cotton production regions in southern Texas, thus reducing time and economic cost for their removal. A GIS network analysis tool was applied to estimate the most likely routes for cotton transportation, and a GIS model was created to identify and visualize potential areas of volunteer cotton growth. The GIS model indicated that, of the 31 counties in southern Texas that may have habitat for volunteer cotton, Hidalgo, Cameron, Nueces, and San Patricio are the counties at the greatest risk. Moreover, a method based on unmanned aerial vehicle (UAV) remote sensing was proposed to detect the precise locations of volunteer cotton plants in potential areas for their subsequent removal. In this study, a UAV was used to scan limited samples of potential volunteer cotton growth areas identified with the GIS model. The results indicated that UAV remote sensing coupled with the proposed image analysis methods could accurately identify the precise locations of volunteer cotton and could potentially assist in the elimination of volunteer cotton along transport routes.

1. Introduction

Large scale cotton production dates back to as early as the 1820s in Texas, which has become a major cotton-producing area in the world (White, 1957) and currently produces nearly half of the U.S. crop. A serious concern for cotton production in the southern part of the state is the encroachment of the boll weevil from Mexico, a major pest that can devastate the crop (Howard, 1903). Boll weevil has been eradicated in almost all of the U.S. but not in Mexico, so cotton plants in southern Texas must be treated for the pest, and cotton plants growing in unmanaged locations must be eliminated to minimize the habitat for boll weevils. Because the seed is still attached to the fiber before the ginning process, seed cotton that falls to the ground can germinate. The resulting unmanaged cotton plants are referred to as volunteer (or feral) cotton, which can potentially provide a host for cotton boll weevils to proliferate.

Volunteer cotton plants can occur in three principal locations. First, they can regrow in fields where a cotton crop was planted in a previous season and seed cotton fell to the ground during the growing season or at harvest. Second, they can grow in areas where unginned cotton was stored at the edge of a field and pieces of seed cotton fell and remained on the ground at that location. Most modern-day cotton harvesters produce round high-density modules of seed cotton that are convenient for harvest logistics as well as storage and transportation. The periphery of the modules is wrapped in plastic to hold picked cotton together and protect it from weathering, but the flat ends are exposed, potentially

* Corresponding authors.
E-mail addresses: timothywangty@tamu.edu (T. Wang), hanxiongzhe@kangwon.ac.kr (X. Han).

https://doi.org/10.1016/j.compag.2021.106629
Received 20 January 2021; Received in revised form 3 December 2021; Accepted 11 December 2021
Available online 27 December 2021

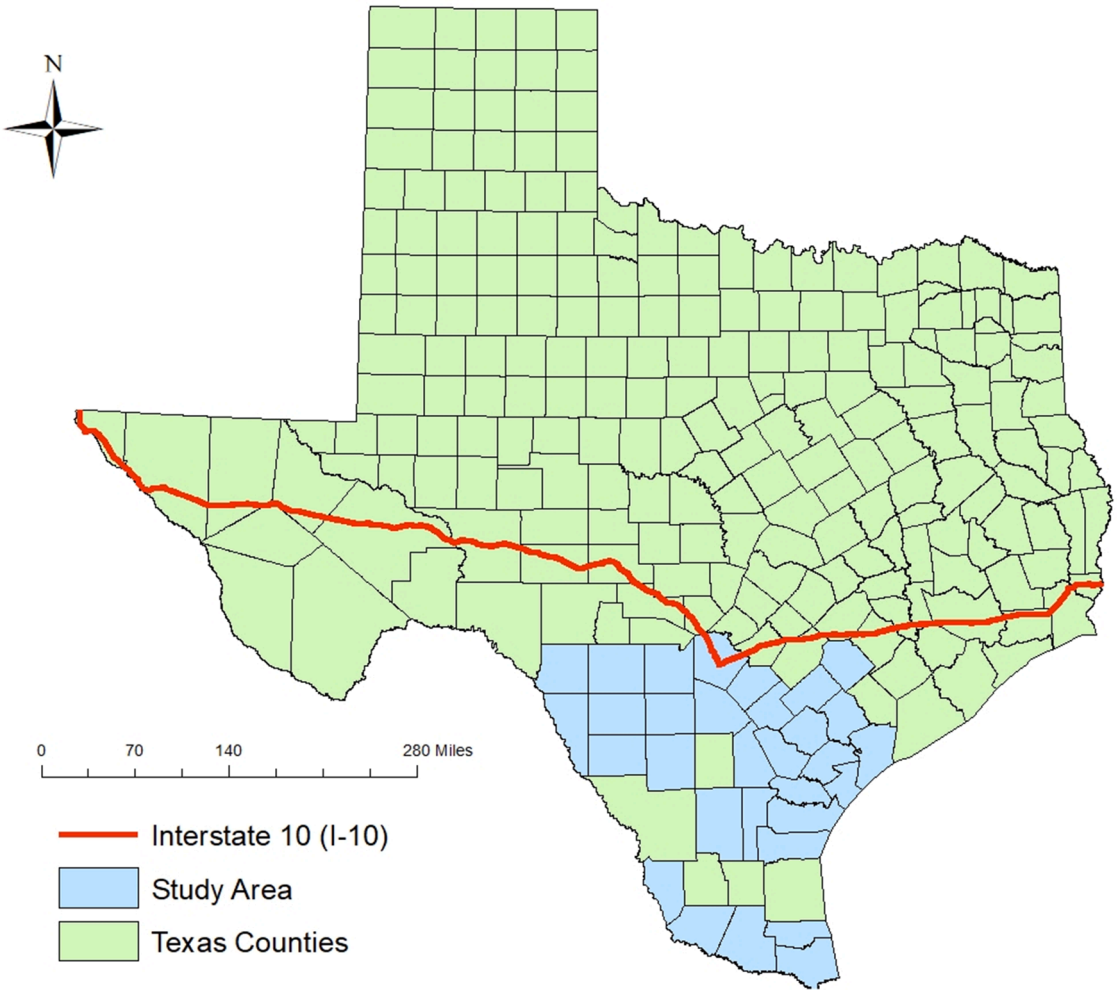

Fig. 1. Interstate 10 (I-10) and the cotton-producing counties in South Texas.

allowing a few pieces of seed cotton to fall to the ground at storage locations. Third, transport of the modules from the field to the ginnery is often done with trucks having no covering, so small pieces of seed cotton from the exposed ends can be deposited along the route, resulting in cotton plants along the roadside. While the first two types of volunteer cotton locations are reasonably well defined, plants along the roadside can be spread over vast areas. To understand the potential extent of volunteer cotton along roadsides, it is necessary to map the common transport routes.

The applications of geographic information systems (GIS) in agriculture have grown since the early days of GIS (Pierce and Clay, 2007), but network analysis, a common GIS tool, has not been widely applied in agriculture. Network analysis calculates the optimum path in a network (Sadeghi-Niaraki et al., 2011), and it has mainly found application in transportation planning. For instance, Duran-Fernandez and Santos (2014) built a GIS model of the National Road Network in Mexico that enables calculation of the best route between any two locations. Sadeghi-Niaraki et al. (2011) proposed a model for route planning that considers factors like weather, sight-seeing information, and road type. Ahmadzai et al. (2019) used network analysis to assess the suitability of urban expansion. Abousaeidi et al. (2016) applied network analysis to create a GIS model that determines the quickest routes for vegetable delivery in Malaysia. The current study aimed to use network analysis and a few additional simple geoprocessing tools to identify potential volunteer cotton growth zones along roadsides in the critical area of southern Texas. If these zones can be identified, some means to detect and remove the volunteer cotton in these zones will be required.

Remote sensing (RS) has been widely used in agricultural production (Khanal et al., 2017; Lee et al., 2010; Read et al., 2003; Sui et al., 2008; Yang, 2012) and shown to be appropriate for identifying individual cotton plants in areas of mixed ground cover (Westbrook et al., 2016). RS with unmanned aerial vehicles (UAVs) is increasingly used in agriculture for applications like disease detection, yield prediction, and production management (Herrmann et al., 2020; Su et al., 2018; Zhang et al., 2019). Previous research has also indicated that UAV RS can be used to identify volunteer cotton in a field where another crop is concurrently growing (Yadav et al., 2019). UAV RS is able to provide image data of high enough resolution to enable cotton plant identification at the single-plant level (Wang et al., 2020a). Multispectral information, including vegetation indices (VIs) like normalized difference vegetation index (NDVI), can be used to distinguish cotton plant pixels from ground pixels because of the difference in reflectance (Wang et al., 2020a,b). However, volunteer cotton plants grown on the roadside are commonly surrounded by other plants with similar reflectance to cotton plants. A major challenge of this research was to differentiate cotton plants from other plants accurately. The height difference between cotton plants and the surrounding plants could contribute to the differentiation.

The canopy height model (CHM) has been commonly used in forestry applications, such as in treetop detection and tree parameter measurement (Guerra-Hernández et al., 2017; Nie et al., 2019; Tian et al., 2019). CHM has also been used to estimate the height of crop plants and to help separate crop rows from soil (Matese et al., 2017). CHM can be derived from light detection and ranging (Lidar) data as well as

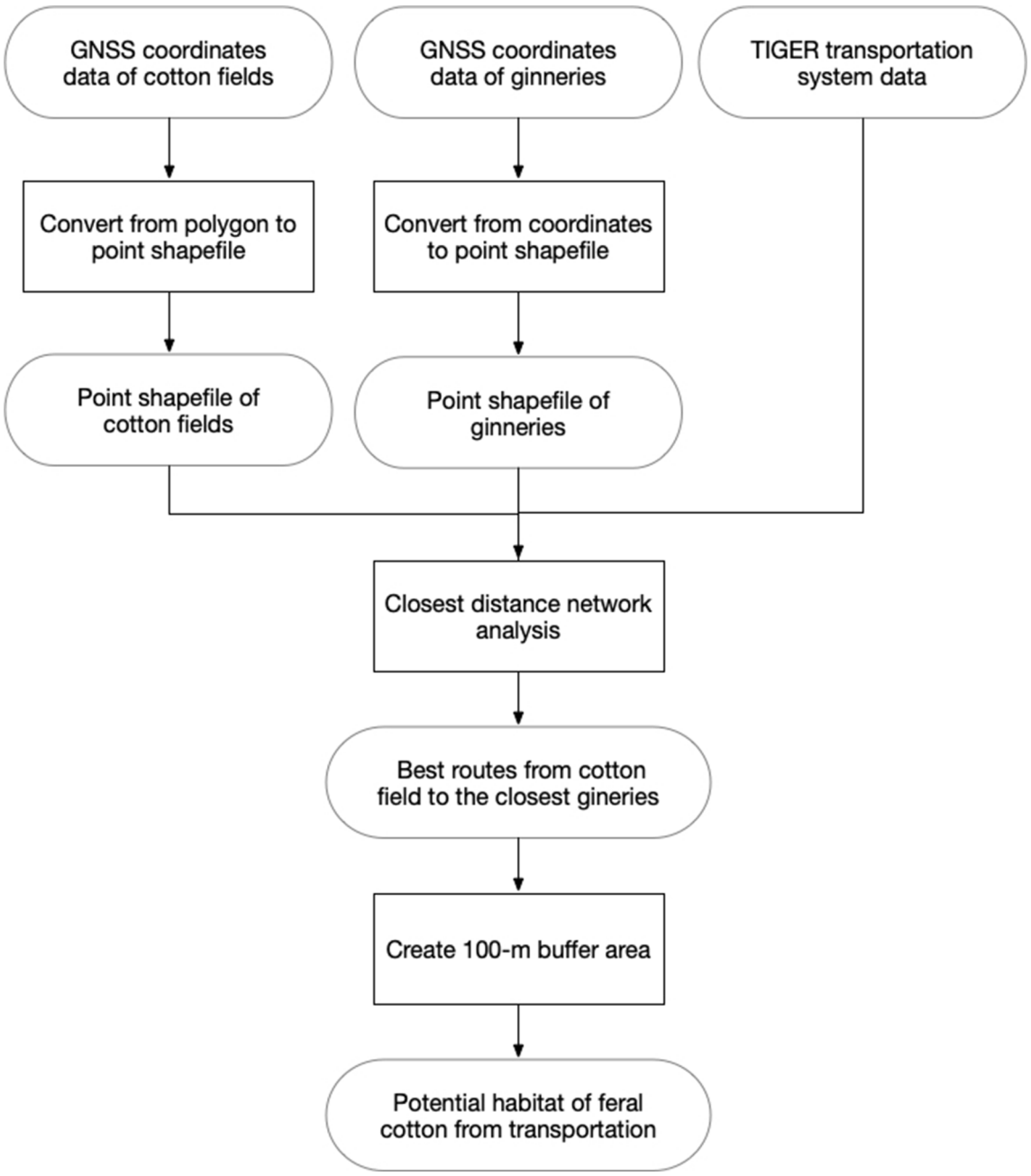

Fig. 2. Flow chart of volunteer cotton habitat prediction model setup.

photogrammetric processing of mosaicked image data (Gu et al., 2020; Han et al., 2018; Jayathunga et al., 2018; Lisein et al., 2013). Because of the rapidly increasing use of UAVs in recent years and the low cost of multispectral sensors relative to Lidar sensors, the photogrammetrically derived CHM has become more common in RS studies. By incorporating high-resolution CHM data, identification of individual volunteer cotton plants may be possible with the help of an object classification algorithm accounting for spectral and height information.

Superpixel segmentation is an image processing algorithm that aggregates the raw pixels of an image to a larger size, superpixels, based on spectral and spatial data. The digital number (DN) values of raw pixels are averaged and assigned as the DN of the superpixel. The method is thus a means of descaling, but it keeps more of the underlying features of an object compared to simple morphological resampling. Therefore, Superpixel segmentation has been used in object detection (Sultani et al., 2017) and tree delineation (Wang et al., 2018), and it has also proven effective for crop disease detection (Zhang et al., 2017, 2018).

Up to now, no prior research has provided an efficient way to locate likely volunteer cotton habitat along roadsides or to identify individual volunteer cotton plants. The goal of this study was thus to provide a means for effective and efficient identification of volunteer cotton plants that might result from transportation in southern Texas cotton production regions. Specific objectives were (1) to develop a GIS-based network analysis model to predict high-potential zones for volunteer cotton habitat resulting from transportation from cotton fields to ginneries, and (2) to develop a method to use UAV RS and image analysis to identify individual volunteer cotton plants for treatment or removal.

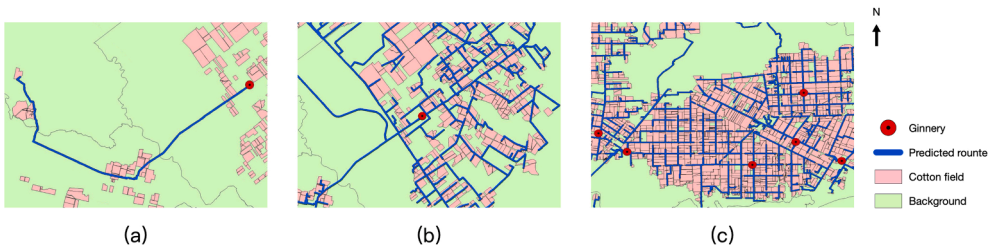

Fig. 3. The basic type of network analysis methods in this study. (a) one to one; (b) many to one; (c) many to many. (Scale 1:500,000).

2. Materials and methods

2.1. Study area

This study focused on cotton fields located to the south of Interstate Highway 10 (I-10; Fig. 1) as possible sources of volunteer cotton. These fields spread across 31 counties in southern Texas with a total area of 8.9 M ha (22 M acres). While UAV RS has the potential for the identification of volunteer cotton plants, monitoring such a large area with UAVs would be unjustifiably costly and time-consuming. Therefore, UAV RS was used in only a few selected areas for detecting volunteer cotton in this study.

2.2. GIS data collection

Historical location information on cotton field locations in southern Texas was provided by the Texas Boll Weevil Eradication Foundation as a digitized, spatially referenced, polygon shapefile. Each polygon in this file represents the shape and location of an individual cotton field. Locations of the ginneries in this region were provided by the Texas Cotton Ginners Association. The road network data for the study area were obtained from the TIGER (Topologically Integrated Geographic Encoding and Referencing) database.

ArcMap 10.5 GIS software (ESRI Corp., Redlands, CA, USA) was used to process the geospatial data and set up the model for analysis. ArcGIS has basic mapping capability as well as advanced geoprocessing capabilities, enabling users to import, visualize, and execute customized instructions with geospatial data (Abousaeidi et al., 2016).

2.3. GIS model setup

A model was developed from the collected data to demarcate the potential habitat area of volunteer cotton along roadsides south of I-10. The data on locations of cotton fields, ginneries, and roads were processed in multiple steps with the network analysis tool (Fig. 2). A network model uses links to represent the linear channels of flow and nodes to represent the origin and destination (Lupien et al., 1987). The coordinates of the ginnery locations were converted into a point shapefile and imported to the network analysis tool. The cotton field boundaries were converted from polygons to points with a data conversion tool in ArcGIS; each cotton field was thus represented by a point located at the centroid of the original polygon. Considering the magnitudes of total transportation distances, the error associated with the distance between field centroids and field edges is negligible.

The first phase of network analysis was to create a geodatabase containing all routes in the study area. The road network data of the 31 counties were thus imported as a feature dataset. The methods of network analysis can be divided into three main categories, one-to-one, many-to-one (or one-to-many), and many-to-many (Fig. 3). One-to-one is used to calculate the optimal route from an individual cotton field to a selected ginnery, whereas many-to-one calculates optimal routes from multiple fields to a ginnery, and many-to-many calculates optimal routes from multiple fields to multiple ginneries simultaneously. The 'closed facilities network analysis' method was applied, which includes characteristics from both the many-to-one and many-to-many approaches. For example, if only one ginnery exists in a particular small area, the best routes between every cotton field in the nearby area and this ginnery are calculated. If two or more ginneries are in the nearby area, the routes between each field and each ginnery are calculated, and the algorithm automatically selects the shortest route among them.

The objective of a network model is to determine the minimum-cost path between the origin and destination, which could either be the shortest distance, the shortest travel time, or a combination of multiple criteria. In this study it was assumed that drivers would select the shortest route for cotton transportation. The road network data used in this study were included in a simple line shapefile that didn't contain specific attribute information of each road segment. Thus, the model assumed all road segments to be bidirectional and ignored possible barriers or turning restrictions. The model ignored detailed factors such as traffic, road conditions, and the speed limits of all road segments. Furthermore, the model did not consider the situations in which cotton farmers may choose the facility to gin their cotton for additional business-related reasons.

Once the model identified the "optimal" routes, a 100-m buffer was

Fig. 4. The images were collected with (a) a DJI Matrice 100 UAV with (b) a Zenmuse X3 NDVI camera.

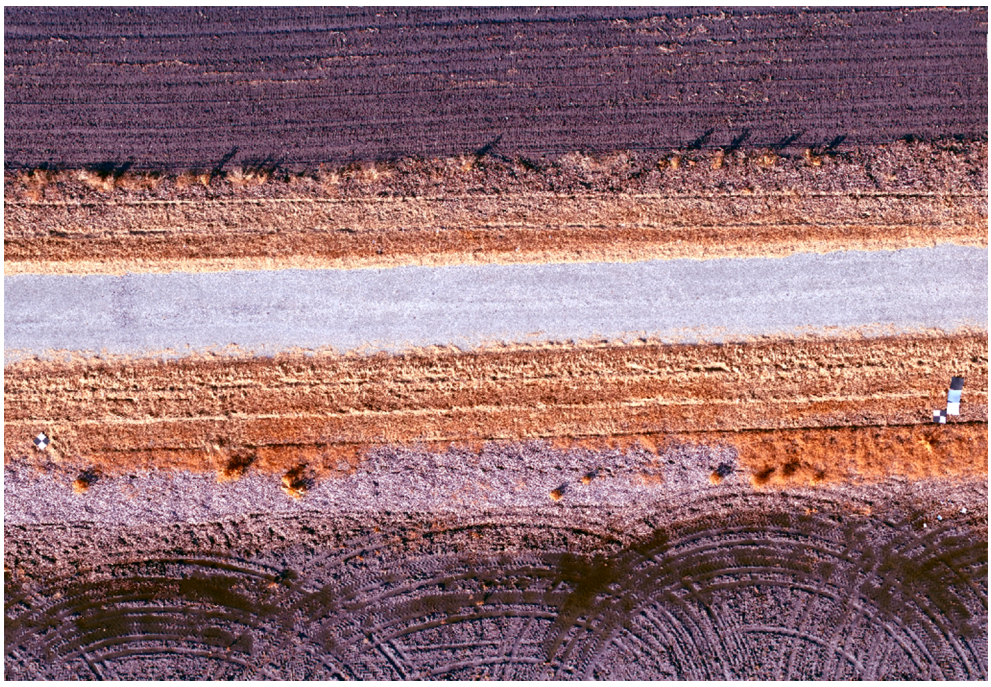

Fig. 5. The false-color image mosaic of an area along a road near Portland, TX (Nov. 16, 2017).

subsequently applied to the routes to estimate potential volunteer cotton growth areas along roadsides. Based on consultation with a an expert in the occurrence volunteer cotton (Yang, 2017), the 100-m buffer was selected as a reasonable maximum distance that seed cotton might be blown from the roadside into surrounding fields and pastures. The areas of potential volunteer cotton habitat along roadsides were calculated

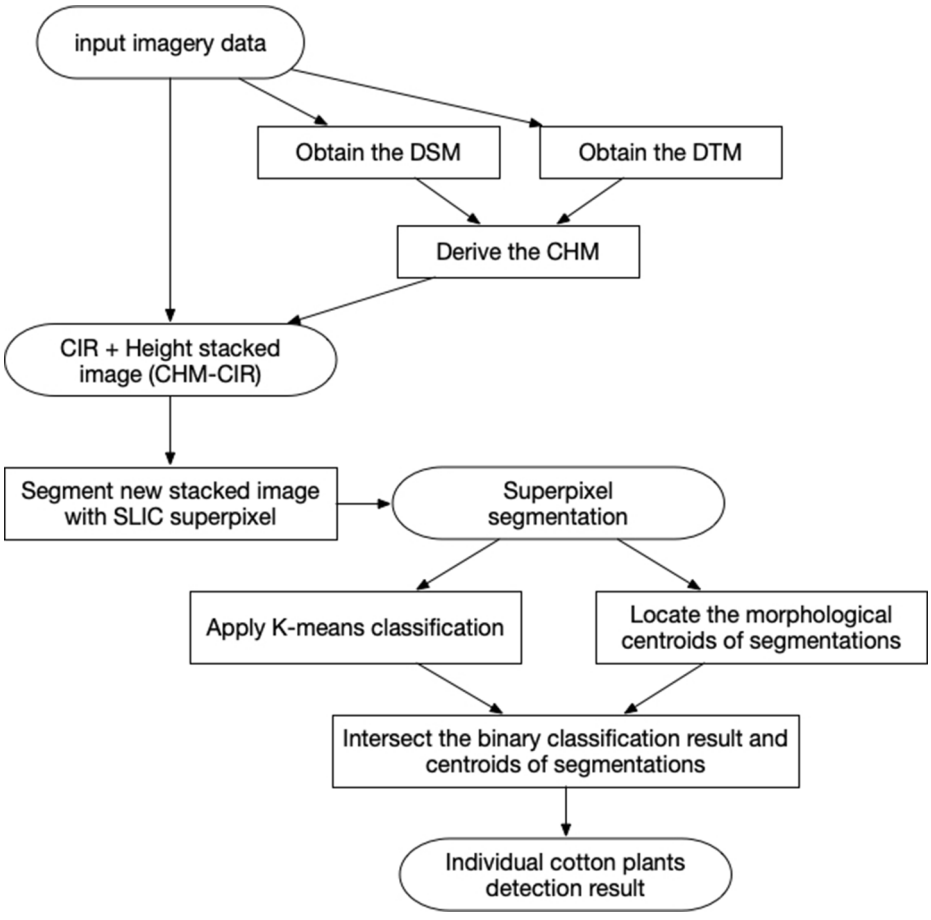

Fig. 6. The flowchart of the plant-level volunteer cotton detection method.

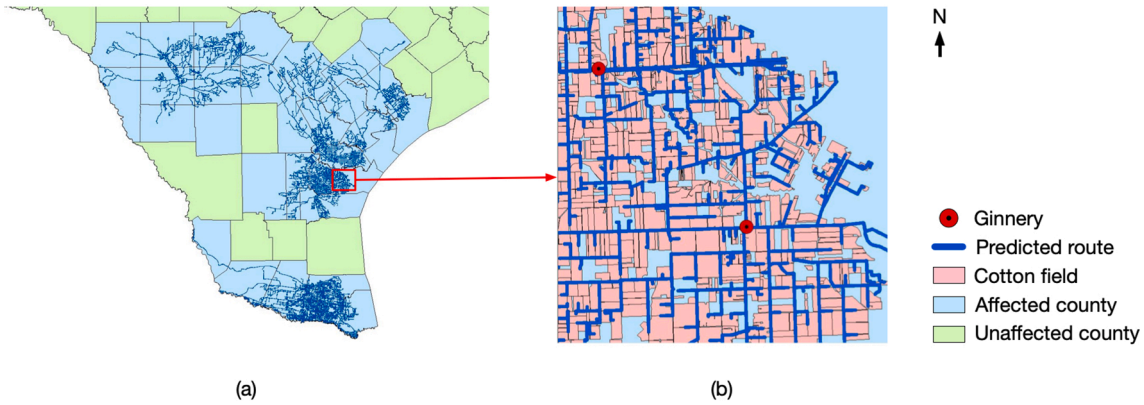

Fig. 7. A GIS map indicating (a) all the habitats of volunteer cotton in southern Texas (Scale 1:7,000,000), and (b) Nueces county in greater detail (Scale 1:3,500,000).

with the network analysis model for each Texas county south of I-10.

2.4. Validation of GIS model and classification based on UAV remote sensing

2.4.1. UAV image data collection

To validate the GIS model, UAV multispectral images were acquired near Portland, Texas, along roadsides within the volunteer-cotton habitat zone predicted by the model on November 16, 2017. Volunteer cotton plants were observed at a specific roadside location, and images were collected with a Zenmuse X3 NDVI camera (DJI Technology Inc., Shenzhen, China; Fig. 4b) mounted on a DJI Matrice 100 rotary-wing UAV (DJI Technology Inc., Shenzhen, China; Fig. 4a). Images of 1280 × 960 pixels were collected at 15.2 m (50 ft) above ground level (AGL), providing for 0.783 cm pixel resolution. The flying speed was around 0.9 m/s with 70% of forward and sideward overlap. The survey covered 0.55 ha in 270 s. Each image contained two bands of data: blue (420–476 nm) and near-infrared (NIR, 764–840 nm). Of the images collected, 135 were mosaicked with Pix4Dmapper software (Pix4D Corp., Lausanne, Switzerland) to generate an orthomosaic for a short segment of road and adjacent roadside area (Fig. 5). The image mosaic was geo-referenced by using three ground control points (GCPs). Three radiometric references (3%, 20%, and 45% reflectance) were used for radiometric calibration.

To reinforce the original findings with the prediction model, an additional round of UAV-based validation was conducted near Woodsboro, Texas, on September 9, 2021. Images were collected with a DJI Matrice 100 carrying a MicaSense RedEdge camera (Micasense, Seattle, WA, USA). The UAV was flying at 30 m AGL with 2.8 m/s of flying speed and 70% of forward and sideward overlap.

Table 1
The volunteer cotton potential habitat area and the potential habitat proportion of each county.

Ranking by area	County name	Potential habitat area (ha)	Potential habitat proportion of county area (%)
1	Hidalgo	43056.57	10.47
2	Cameron	34347.44	10.74
3	Nueces	33188.06	10.82
4	San Patricio	23622.13	12.90
5	Medina	20892.16	6.05
6	Willacy	19028.62	9.54
7	Jim Wells	16806.57	7.48
8	Uvalde	15176.17	3.76
9	Zavala	13364.11	3.98
10	Frio	13353.99	4.57
11	Bee	12609.81	5.54
12	Victoria	11392.40	4.96
13	Kleberg	10558.06	3.75
14	Refugio	10477.93	5.03
15	Karnes	10419.33	5.37
16	Atascosa	9878.02	3.14
17	Calhoun	7511.50	2.61
18	Starr	7426.47	2.33
19	Wilson	7022.44	3.37
20	Live Oak	6681.41	2.40
21	DeWitt	5890.57	2.51
22	Goliad	5058.98	2.28
23	Bexar	3797.09	1.17
24	Kinney	3247.40	0.92
25	Maverick	2588.00	0.78
26	Dimmit	1980.85	0.58
27	Duval	1956.06	0.42
28	Zapata	1780.74	0.65
29	Lavaca	1683.80	0.67
30	Aransas	339.27	0.23
31	La Salle	295.37	0.08

2.4.2. Classification and validation

Various types of live vegetation like grass and weeds were observed along the road. Because cotton plants have a bushy architecture and are commonly taller than most of the surrounding plants on the roadside, the height of vegetation was considered along with the spectral information to help distinguish cotton plants from other plants. The height of the vegetation was determined by creating a CHM, derived by subtracting the digital terrain model (DTM) from the digital surface model (DSM). A DTM generated with Pix4D can only achieve pixel resolution in that is five times coarser than the DSM. Therefore, the DTM was upsampled to match the DSM's resolution before being subtracted from the DSM. The DSM and DTM were calculated by means of structure from motion (SfM) processing in Pix4D. The blue and NIR spectral data and height (CHM) data were stacked to form a visual image (referred to as the CHM-CIR image) used in a classification process for volunteer cotton plants (Fig. 6). An algorithm based on simple linear iterative clustering (SLIC) Superpixel segmentation, k-means clustering, and morphological image processing was developed to automatically detect individual volunteer cotton plants with the CHM-CIR image. The CHM-CIR image was segmented into superpixels based on similar spatial (including height) and spectral information. The number of segments was user-determined according to the resolution of the input image and the size of the cotton plants in the image from the canopy view. The DN value of each superpixel was the average of the DN values of the original pixels within the area of the superpixel. A two-class k-means classification was applied to the superpixel image to generate a binary regional classification map, which indicated the locations of volunteer cotton plants. In the binary map, volunteer cotton regions were assigned a value of 1, while all other regions were assigned a value of 0. To improve the precision of identification, a morphological process was applied to the

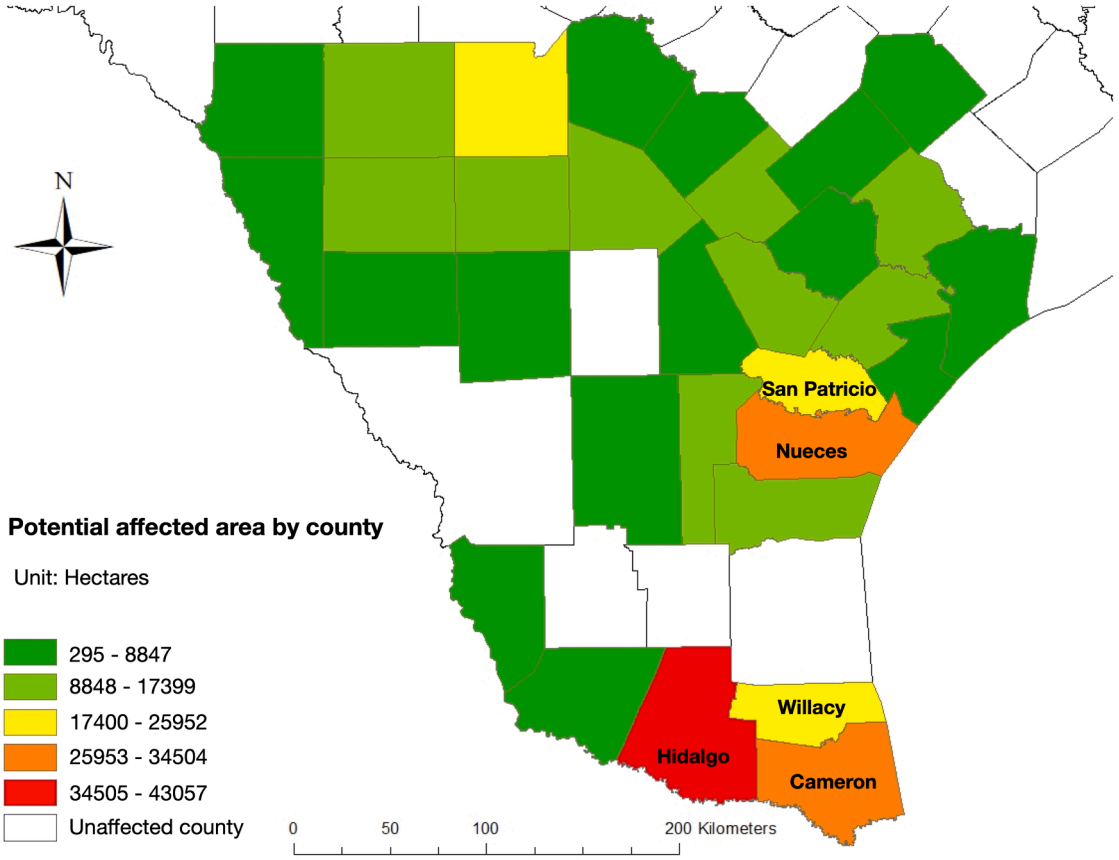

Fig. 8. The risk severity of volunteer cotton caused by transportation shown by counties in southern Texas.

superpixel image to identify the centroids of the superpixels. In the output of this process, centroids were marked with a 1, while non-centroid areas were marked with a 0. The intersection of the binary classified image and the centroid output formed a map of individual cotton plants, with the center points of each cotton plant clearly marked out and given a value of 1.

To validate the classification result, ground-truth data were collected in the study area. Specifically, GPS coordinates of the volunteer cotton plants were collected with a Trimble Geoexplorer 6000 GPS receiver (Trimble, Sunnyvale, CA). A confusion matrix including overall accuracy, kappa coefficient, errors of commission, and errors of omission, was generated for accuracy assessment. The classification was assessed based on whether each superpixel was correctly classified or not. The kappa coefficient represents the agreement between predicted class type and ground-truth class type. The errors of commission represent overclassification of a superpixel into the volunteer cotton plant, while the errors of omission represent underclassification of a superpixel where volunteer cotton existed but was not identified.

3. Results and discussion

3.1. Network analysis

The map of potential volunteer cotton habitat resulting from transporting seed cotton from fields to gins (Fig. 7) shows the potential area that would need to be surveilled across the 31 counties south of I-10. The potential habitat area was 355,431 ha, less than 4% of the total area of the 31 counties.

The potential habitat area in each of these Texas counties was calculated (Table 1), and the proportion in each county ranged from less

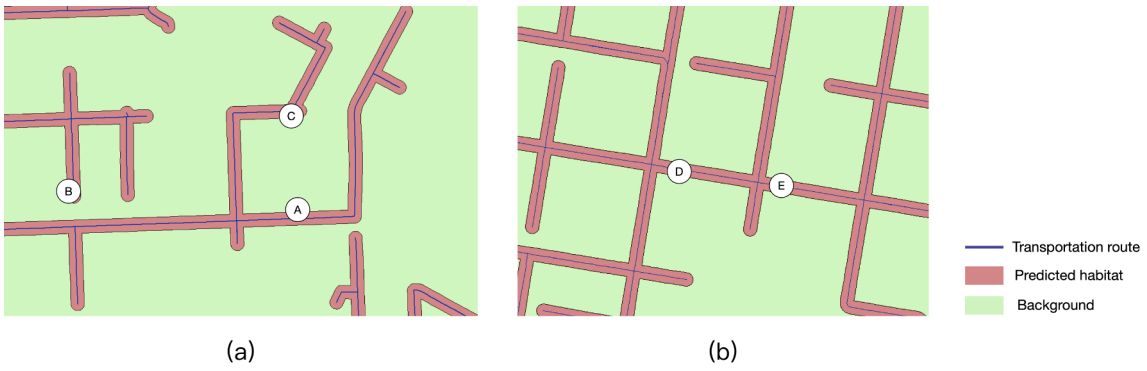

Fig. 9. Three predicted habitat locations in (a) Portland, TX of Nueces county and two predicted habitat locations in (b) Woodsboro, TX of Refugio county validated to have volunteer cotton plants.

Fig. 10. The ground-truth images corresponding to Locations A, B, and C in Fig. 9. All of these cotton plants were located along the roadside.

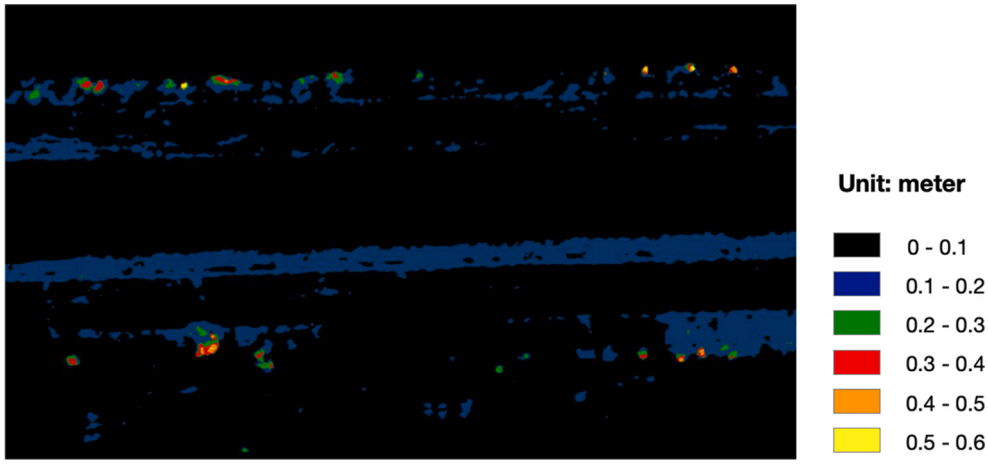

Fig. 11. The CHM image of Location A. The height of objects ranged from 0 to 0.6 m.

than 0.1% to nearly 13%. A map of these counties was colored according to the potential volunteer cotton area (Fig. 8) based on the network analysis model. Hidalgo and Cameron counties had the largest (43,057 ha) and second-largest (34,347 ha) potential volunteer cotton habitat areas, respectively. These counties are located on the border with Mexico and thus have a high risk of encroachment by cotton boll weevils across the Mexican border. Nueces county ranked third with 33,188 ha of potential volunteer cotton area. This county is only 161 km (100 miles) from the border, so it also presents a concern as a potential boll weevil habitat. San Patricio county, adjacent to and north of Nueces County, ranked fourth with 23,475 ha of potential infested habitat area. Nueces and San Patricio counties surround the cotton-intensive Corpus Christi Bay area and had the highest proportions of potential habitat (10.8% and 12.9%, respectively), so they were of particular interest.

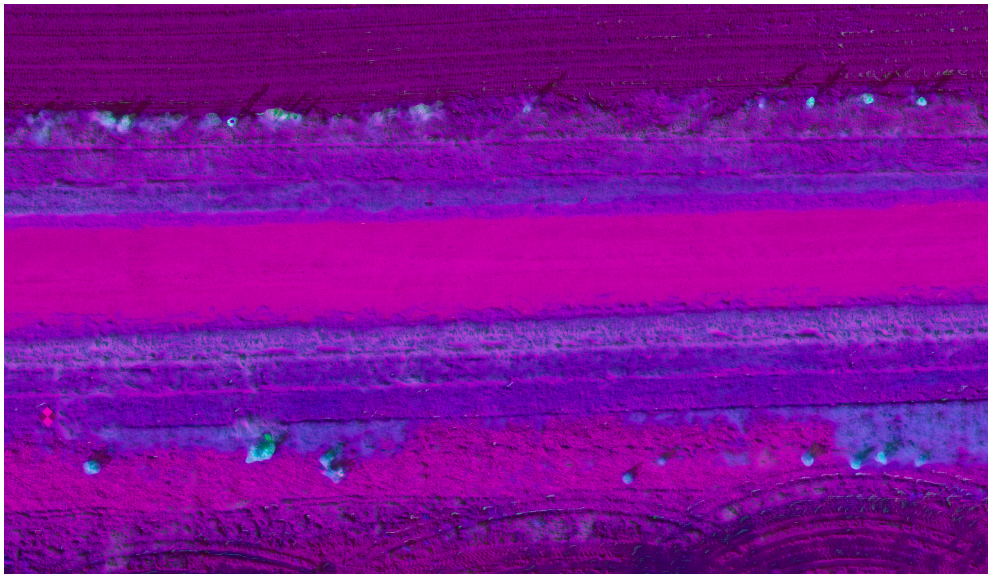

Fig. 12. The composite image of blue, CHM, and NIR (CHM-CIR). The green (cyan) color clusters represent cotton plants. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

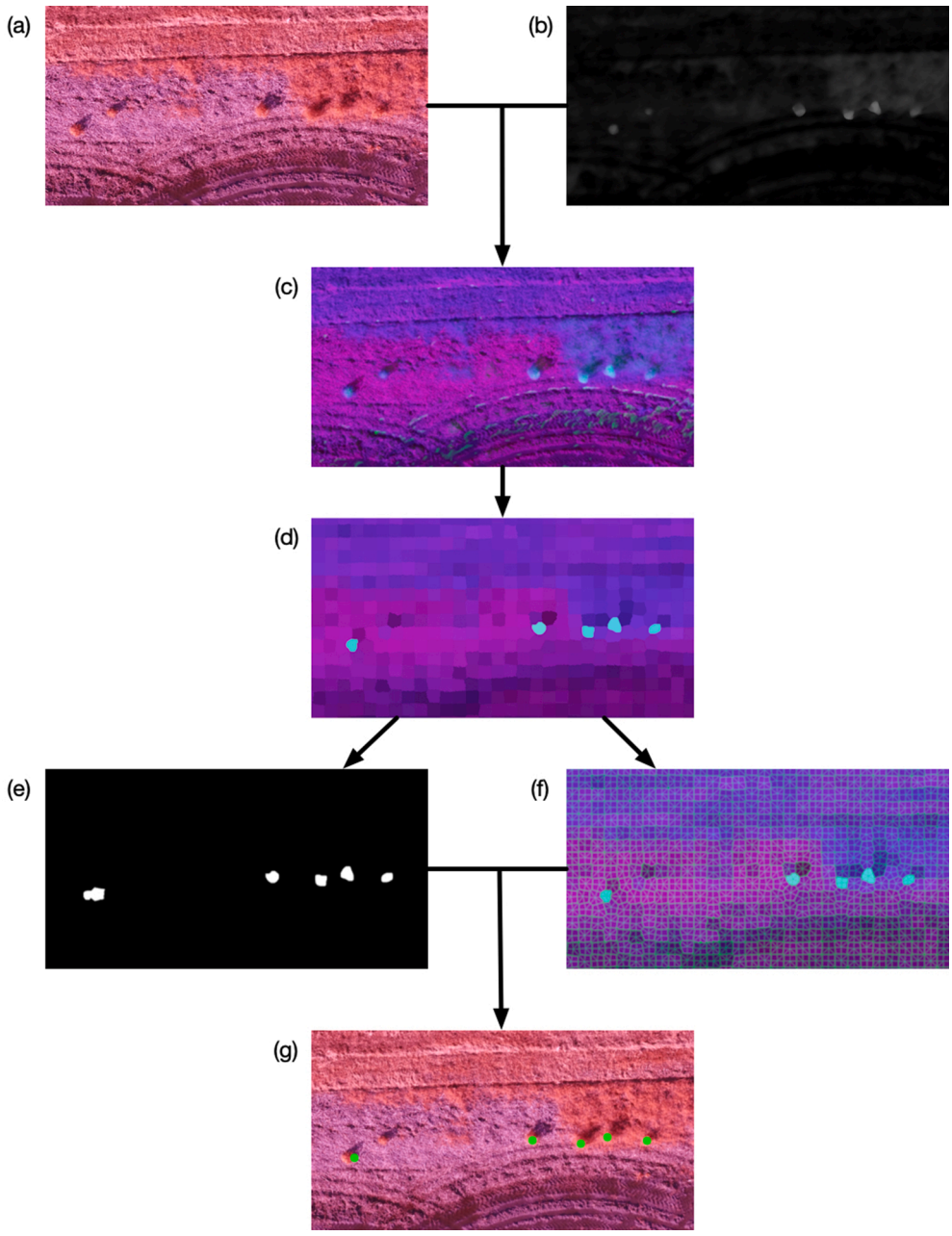

Fig. 13. The image processing results for each step of the volunteer cotton identification algorithm on the example image. (a) Raw multispectral image; (b) CHM data; (c) CHM-CIR image; (d) Superpixel segmentation image; (e) Regional classification for indication volunteer cotton area; (f) Centroid location of each segmentation; and (g) The final identification result of volunteer cotton.

3.2. Selecting remote-sensing locations based on the GIS model

Volunteer cotton plants were ground-truthed at five predicted locations (A, B, C, D and E in Fig. 9) near Portland, Texasin Nueces County and Woodsboro, Texas in Refugio county, where volunteer cotton plants were readily observed. Most of the cotton plants grew along the roadside within roughly 3 to 10 m from the road. At Location B, volunteer cotton plants were found in both small bush and larger tree forms (Fig. 10) more than 50 m from the road, indicating some plants had been growing there for a significant time period. UAV RS was conducted to classify volunteer cotton at Location A. In a CHM image of Location A, the height and location of small bush-form cotton plants (Fig. 11, bottom left) were visible and differentiable from tree-form cotton plants (Fig. 11, top right). The canopy height of the volunteer cotton plants in the study ranged from a few cm to 0.6 m. One volunteer cotton plant was found at each location, D and E.

3.3. Classification and validation

The CHM data were linearly normalized into the value range of 0–255 to match the range of DN values of the blue and NIR bands. The raw output image of the Zenmuse NDVI camera had the blue and NIR bands assigned to blue and red bands for visualization. To minimize changes in the output composite images, CHM was assigned to the green band, while blue and NIR were kept in the original bands (Fig. 12). Green (cyan) color clusters tended to represent tall objects, which were generally cotton plants in this case.

For the purpose of demonstration, the northeast corner of Location A is used as an example of the image composition and classification process in Fig. 13. The raw multispectral image (Fig. 13a) and the CHM data

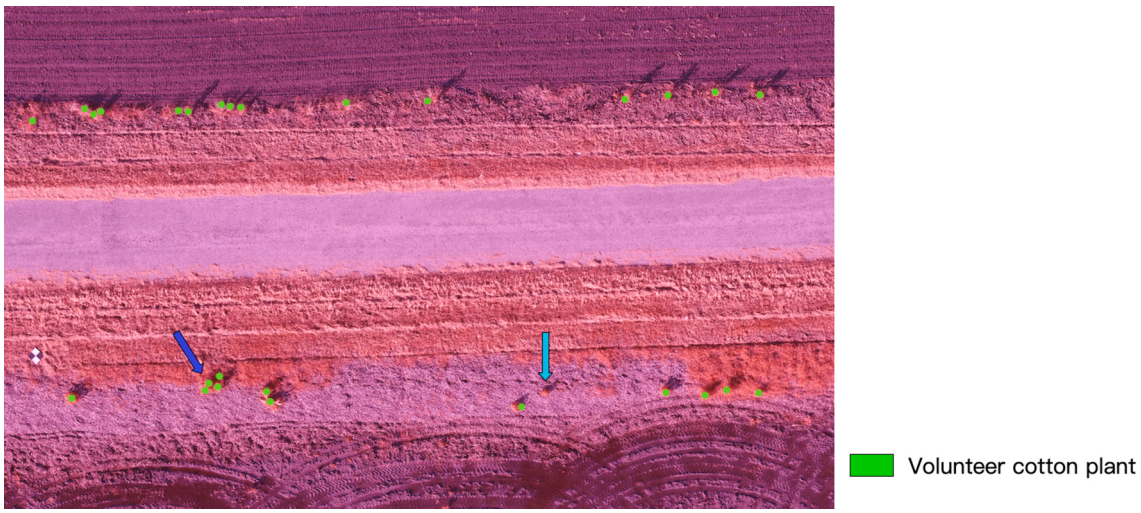

Fig. 14. The plant-level classification result. Each green dot represents an individual volunteer cotton plant. The misclassified and overclassified plants are highlighted with light blue and dark blue arrows, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The accuracy and confusion matrix of plant-level volunteer cotton classification.

Overall accuracy		**99.95%**				
Kappa Coefficient		0.9630				
	Class types determined from reference source (Ground-truth)				Commission	Omission
Class types determined from classified map		Cotton	Others	Totals		
	Cotton	26	1	27	1/27	1/27
					3.70%	3.70%
	Others	1	4,340	4,341	1/4341	1/4341
					0.02%	0.02%
	Totals	27	4,341	4,368		

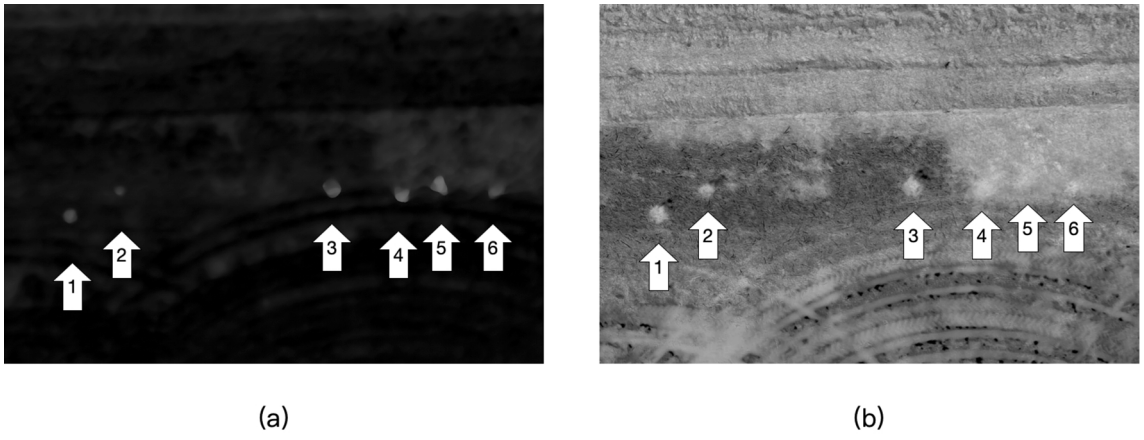

Fig. 15. All six cotton plants are clearly shown in (a) CHM data, while only 3 of 6 can be observed from (b) Blue NDVI data.

(Fig. 13b) were stacked to form the CHM-CIR image (Fig. 13c). SLIC Superpixel segmentation was applied to the CHM-CIR image to get the segmentation image (Fig. 13d). K-means classification was applied to the segmentation image to obtain the two-class classification result (Fig. 13e), while the centroids of superpixels were identified (Fig. 13f) based on the segmentation image. The identified cotton plants were marked with green dots (Fig. 13g).

The plant-level classification result of the entire Location A region is shown in Fig. 14. Volunteer cotton plants were detected correctly in 96.3% of cases (26 of 27). One plant was misclassified (light blue arrow in Fig. 14) and one plant was overclassified (dark blue arrow in Fig. 14).

The image was segmented into 4,368 superpixels (Table 2). The overall accuracy reached 99.95% with a Kappa coefficient of 0.9630. It must be noted, however, that the high accuracy was based on a relatively small area. Thus the commission and omission errrors for cotton, both of which were 3.70% in this particular case, were more meaningful.

The misclassified cotton plant (Fig. 13a, plant #2) was observed to be similar to the rest in terms of shadows and color variance. However, its height was less than that of the other cotton plants, so the CHM data did not suggest that the object was a cotton plant. In future studies, a filter that generates a buffer area around tall objects should be developed to increase the weight of CHM information in image processing.

The Blue-NDVI index (Eq. (1)) was also considered as a potential indicator to identify the cotton plants.

$$Blue\ \ NDVI = \frac{NIR - Blue}{NIR + Blue} \quad (1)$$

However, the reflectance signature of cotton and surrounding plants was similar, resulting in similar Blue-NDVI values (Fig. 15). The CHM data were decisive in identifying all six cotton plants, but Blue-NDVI contributed little to the differentiation of cotton plants #4, #5, and #6. In most cases, spectral and height information are expected to provide adequate discrimination, but when cotton and surrounding plants have similar height and spectral reflectance, the algorithm used here will have difficulty distinguishing them without additional contextual information.

CHM-CIR images have been shown to be able to distinguish cotton plants from the ground, non-vegetation objects, and other types of short plants like short weeds. However, if the weeds are tall, such as well-developed Ragweed or Palmer Amaranth, these methods alone may not be able to differentiate cotton plants from weeds. In general, however, this type of UAV RS data can potentially eliminate the effect of other plants and provide accurate classificationof volunteer cotton plants, or even other types of crop plants needing to be differentiated from their surroundings.

Using UAV RS to identify volunteer cotton is still exploratory, and two directions could be followed to improve the process in the future. If high-resolution images are being used, cotton bolls can potentially be identified as another criterion to differentiate cotton plants. If the resolution is too coarse to identify cotton bolls, morphological and deep learning algorithms may be able to help. Morphological tools such as segmentation can be used to identify cotton plants according to their features like shape and color, etc. At mean time, deep learning has good performance on automatically feature extraction.

UAVs currently have limitations making them not feasible for covering large areas, but future improvement will likely alleviate this problem. For example, issues such as flight time can be overcome with breakthroughs in battery technology and removal of airspace restrictions against flying beyond visual line of sight (BVLOS).

4. Conclusion

Volunteer cotton in southern Texas creates the risk of providing a habitat for destructive cotton boll weevils. In this study, a GIS-based framework for effective and efficient identification of volunteer cotton plants associated with transporting seed cotton from fields to gins was established. The proposed GIS network analysis model drastically reduced the size of potential habitat area requiring surveillance, enabling a focus on transportation routes between cotton fields and cotton ginneries. A total of 31 counties in southern Texas have potential habitat for volunteer cotton, and the proposed GIS model identified Hidalgo, Cameron, Nueces, and San Patricio as particularly high-risk counties for volunteer cotton habitat. Three predicted locations in Nueces county were used as examples to show volunteer cotton along the roadside. A plant-level classification method based on UAV RS was developed. The method was able to identify volunteer cotton plants automatically with excellent overall accuracy. The proposed GIS model and the plant-level volunteer cotton classification method can potentially enable precise and timely treatment and control of volunteer cotton.

CRediT authorship contribution statement

Tianyi Wang: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Visualization, Investigation, Validation. **Xiaohan Mei:** Conceptualization, Methodology, Data curation, Writing – original draft, Investigation, Validation. **J. Alex Thomasson:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition. **Chenghai Yang:** Conceptualization, Writing – review & editing, Resources. **Xiongzhe Han:** Writing – review & editing, Methodology. **Pappu Kumar Yadav:** Data curation, Writing – review & editing, Validation. **Yeyin Shi:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by USDA-APHIS, Research Agreement number 16-8130-0741-CA. The authors thank Texas Boll Weevil Eradication Foundation and Texas Cotton Ginnery Association for providing the GIS data for this study. Thanks are also extended to Cody Bagnall for helping in data collection.

References

Abousaeidi, M., Fauzi, R., Muhamad, R., 2016. Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. Saudi J. Biol. Sci. 23 (5), 555–564. https://doi.org/10.1016/j.sjbs.2015.06.004.

Ahmadzai, F., Rao, K.M.L., Ulfat, S., 2019. Assessment and modelling of urban road networks using Integrated Graph of Natural Road Network (a GIS-based approach). J. Urban Manag. 8 (1), 109–125. https://doi.org/10.1016/j.jum.2018.11.001.

Duran-Fernandez, R., Santos, G., 2014. A GIS model of the National Road Network in Mexico. Res. Transp. Econ. 46, 36–54. https://doi.org/10.1016/j.retrec.2014.09.004.

Gu, J., Grybas, H., Congalton, R.G., 2020. A comparison of forest tree crown delineation from unmanned aerial imagery using canopy height models vs. spectral lightness. Forests 11, 605. https://doi.org/10.3390/f11060605.

Guerra-Hernández, J., González-Ferreiro, E., Monleón, V.J., Faias, S.P., Tomé, M., Díaz-Varela, R.A., 2017. Use of multi-temporal UAV-derived imagery for estimating individual tree growth in Pinus pinea stands. Forests 8, 1–20. https://doi.org/10.3390/f8080300.

Han, X., Thomasson, J.A., Bagnall, G.C., Pugh, N.A., Horne, D.W., Rooney, W.L., Jung, J., Chang, A., Malambo, L., Popescu, S.C., Gates, I.T., Cope, D.A., 2018. Measurement and Calibration of Plant-Height from Fixed-Wing UAV Images. Sensors 18, 4092. https://doi.org/10.3390/s18124092.

Herrmann, I., Bdolach, E., Montekyo, Y., Rachmilevitch, S., Townsend, P.A., Karnieli, A., 2020. Assessment of maize yield and phenology by drone-mounted superspectral camera. Precis. Agric. 21 (1), 51–76. https://doi.org/10.1007/s11119-019-09659-5.

Howard, L.O., 1903. The Mexican Cotton Boll Weevil. Science 18 (465), 693.

Jayathunga, S., Owari, T., Tsuyuki, S., 2018. Evaluating the performance of photogrammetric products using fixed-wing UAV imagery over a mixed conifer-broadleaf forest: Comparison with airborne laser scanning. Remote Sens. 10, 187. https://doi.org/10.3390/rs10020187.

Khanal, S., Fulton, J., Shearer, S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. Comput. Electron. Agric. 139, 22–32. https://doi.org/10.1016/j.compag.2017.05.001.

Lee, W.S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., Li, C., 2010. Sensing technologies for precision specialty crop production. Comput. Electron. Agric. 74 (1), 2–33. https://doi.org/10.1016/j.compag.2010.08.005.

Lisein, J., Pierrot-Deseilligny, M., Bonnet, S., Lejeune, P., 2013. A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery. Forests 4, 922–944. https://doi.org/10.3390/f4040922.

Lupien, A.E., Moreland, W.H., Dangermond, J., 1987. Network analysis in geographic information systems. Photogramm. Eng. Remote Sens. 53, 1417–1421.

Matese, A., Di Gennaro, S.F., Berton, A., 2017. Assessment of a canopy height model (CHM) in a vineyard using UAV-based multispectral imaging. Int. J. Remote Sens. 38 (8-10), 2150–2160. https://doi.org/10.1080/01431161.2016.1226002.

Nie, S., Wang, C., Xi, X., Luo, S., Zhu, X., Li, G., Liu, H., Tian, J., Zhang, S.u., 2019. Assessing the impacts of various factors on treetop detection using LiDAR-derived canopy height models. IEEE Trans. Geosci. Remote Sens. 57 (12), 10099–10115. https://doi.org/10.1109/TGRS.2019.2931408.

Pierce, F.J., Clay, D., 2007. GIS applications in agriculture. CRC Press.

Read, J.J., Iqbal, J., Thomasson, J.A., Willers, J.L., Jenkins, J.N., 2003. Remote sensing in dryland cotton: Relation to yield potential and soil properties. Ecosyst. Dyn. Agric. Remote Sens. Model. Site-Specific Agric. 5153, 61–72. https://doi.org/10.1117/12.505875.

Sadeghi-Niaraki, A., Varshosaz, M., Kim, K., Jung, J.J., 2011. Real world representation of a road network for route planning in GIS. Expert Syst. Appl. 38 (10), 11999–12008. https://doi.org/10.1016/j.eswa.2010.12.123.

Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., Li, Q., Guo, L., Chen, W., 2018. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. Comput. Electron. Agric. 155, 157–166. https://doi.org/10.1016/j.compag.2018.10.017.

Sui, R., Thomasson, J.A., Hanks, J., Wooten, J., 2008. Ground-based sensing system for weed mapping in cotton. Comput. Electron. Agric. 60 (1), 31–38. https://doi.org/10.1016/j.compag.2007.06.002.

Sultani, W., Mokhtari, S., Yun, H.-B., 2017. Automatic pavement object detection using Superpixel segmentation combined with conditional random field. IEEE Trans. Intell. Transp. Syst. 19 (7), 2076–2085. https://doi.org/10.1109/TITS.2017.2728680.

Tian, J., Dai, T., Li, H., Liao, C., Teng, W., Hu, Q., Ma, W., Xu, Y., 2019. A novel tree height extraction approach for individual trees by combining TLS and UAV image-based point cloud integration. Forests 10, 1–18. https://doi.org/10.3390/f10070537.

Wang, M.i., Dong, Z., Cheng, Y., Li, D., 2018. Optimal segmentation of high-resolution remote sensing image by combining superpixels with the minimum spanning tree. IEEE Trans. Geosci. Remote Sens. 56 (1), 228–238. https://doi.org/10.1109/TGRS.2017.2745507.

Wang, T., Thomasson, J.A., Isakeit, T., Yang, C., Nichols, R.L., 2020a. A Plant-by-Plant Method to Identify and Treat Cotton Root Rot Based on UAV Remote Sensing. Remote Sens. 12, 2453. https://doi.org/10.3390/rs12152453.

Wang, T., Thomasson, J.A., Yang, C., Isakeit, T., Nichols, R.L., 2020b. Automatic Classification of Cotton Root Rot Disease Based on UAV Remote Sensing. Remote Sens. 12, 1310. https://doi.org/10.3390/rs12081310.

Westbrook, J.K., Eyster, R.S., Yang, C., Suh, C.P.C., 2016. Airborne multispectral identification of individual cotton plants using consumer-grade cameras. Remote Sens. Appl. Soc. Environ. 4, 37–43. https://doi.org/10.1016/j.rsase.2016.02.002.

White, R.E., 1957. Cotton ginning in Texas to 1861 61, 257–269.

Yadav, P., Thomasson, J.A., Enciso, J., Samanta, S., Shrestha, A., 2019. Assessment of different image enhancement and classification techniques in detection of volunteer cotton using UAV remote sensing. In: Proceedings of SPIE. p. 20. https://doi.org/10.1117/12.2518721.

Yang, C., 2017. Personal communication on volunteer cotton distribution. USDA-ARS.

Yang, C., 2012. A high-resolution airborne four-camera imaging system for agricultural remote sensing. Comput. Electron. Agric. 88, 13–24. https://doi.org/10.1016/j.compag.2012.07.003.

Zhang, L., Zhang, H., Niu, Y., Han, W., 2019. Mapping maizewater stress based on UAV multispectral remote sensing. Remote Sens. 11, 605. https://doi.org/10.3390/rs11060605.

Zhang, S., Wang, H., Huang, W., You, Z., 2018. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. Optik (Stuttg) 157, 866–872. https://doi.org/10.1016/j.ijleo.2017.11.190.

Zhang, S., Zhu, Y., You, Z., Wu, X., 2017. Fusion of superpixel, expectation maximization and PHOG for recognizing cucumber diseases. Comput. Electron. Agric. 140, 338–347. https://doi.org/10.1016/j.compag.2017.06.016.