

Machine learning-based crop recognition from aerial remote sensing imagery

Yanqin TIAN¹, Chenghai YANG (✉)², Wenjiang HUANG¹, Jia TANG¹, Xingrong LI¹, Qing ZHANG (✉)¹

¹ Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
² USDA-Agricultural Research Service, Aerial Application Technology Research Unit, College Station, TX 77845, USA

© Higher Education Press 2021

Abstract Timely and accurate acquisition of crop distribution and planting area information is important for making agricultural planning and management decisions. This study employed aerial imagery as a data source and machine learning as a classification tool to statically and dynamically identify crops over an agricultural cropping area. Comparative analysis of pixel-based and object-based classifications was performed and classification results were further refined based on three types of object features (layer spectral, geometry, and texture). Static recognition using layer spectral features had the highest accuracy of 75.4% in object-based classification, and dynamic recognition had the highest accuracy of 88.0% in object-based classification based on layer spectral and geometry features. Dynamic identification could not only attenuate the effects of variations on planting dates and plant growth conditions on the results, but also amplify the differences between different features. Object-based classification produced better results than pixel-based classification, and the three feature sets (layer spectral alone, layer spectral and geometry, and all three) resulted in only small differences in accuracy in object-based classification. Dynamic recognition combined with object-based classification using layer spectral and geometry features could effectively improve crop classification accuracy with high resolution aerial imagery. The methodologies and results from this study should provide practical guidance for crop identification and other agricultural mapping applications.

Keywords machine learning, crop recognition, aerial imagery, dynamic recognition, static recognition

1 Introduction

Food security (Mutanga et al., 2017; Farg et al., 2019) and sustainable agricultural development (Kenduiywo et al., 2016) are hot topics in today's society. In this social context, information such as soil fertility, farming systems, disaster monitoring, spatial distribution of crops, and planting area all directly or indirectly affect every management decision of the decision maker. Remote sensing technology enables the acquisition and digitization of these data (Murmu and Biswas, 2015; Wu et al., 2018; Damian et al., 2020), and can provide cross-scale spatial and temporal consistency information (Dimov et al., 2019). As early as in 1972, the Earth Resources Satellite (ERTS) Multispectral Scanner (MSS) sensor was used to acquire remote sensing images and applied to crop identification (Bauer and Cipra, 1973). Since then, more and more commercial satellites and customized airborne imaging systems have been developed and applied to agricultural remote sensing (Zhang et al., 2016). However, due to the complexity of environmental and climatic conditions, there are still great challenges in remote sensing mapping (Wu et al., 2014; Zhang et al., 2014). It is difficult for general producers and managers to master the data processing technology of different remote sensing systems. At the same time, the low spatial resolution of traditional satellite imagery is not suitable for precision agricultural applications (Torres-Sánchez et al., 2014). With the advantages of low cost, high spatial resolution, compact data storage, and ease of use, consumer-grade digital cameras have become a popular tool in remote sensing applications (Yang and Hoffmann, 2015; Song et al., 2016), especially for agricultural remote sensing (Sakamoto et al., 2012). These cameras can be easily carried on manned aircraft and low-flying platforms such as drones and hot air balloons, further accelerating their applications.

Machine learning, which can identify patterns and

Received September 3, 2020; accepted December 7, 2020

E-mails: chenghai.yang@usda.gov (Chenghai YANG);
zhangqing202017@aircas.ac.cn (Qing ZHANG)

correlations and discover knowledge from data sets, is the core of artificial intelligence, focusing on learning (Van Klompenburg et al., 2020). In general, machine learning is about learning patterns from a large amount of historical data through relevant algorithms, making predictions or judgments about new sample data, and then learning like a human (Wang et al., 2021). It has been applied to various data prediction applications, such as image processing, natural language processing and so on (Cai et al., 2018). The biggest difference between non-parametric and parametric machine learning algorithms is that non-parametric algorithms do not assume that the data are normally distributed for a particular image. This makes non-parametric machine learning algorithms more flexible and can be used to predict high-performance models (Zheng et al., 2015). For example, decision tree (Zhang et al., 2016) and support vector machine (Pal and Foody, 2010) have good performance in crop recognition.

With the rapid development of remote sensing systems, image classification technology for crop identification is also constantly improving. There are two main types of classification methods for crop identification: static identification and dynamic identification. Static identification is to select a single remote sensing image with more complete vegetation characteristics as a data source (Van Niel and McVicar, 2004; Boryan et al., 2011; Yang et al., 2011). Dynamic recognition uses multiple periods or time series (usually the entire growing season) remote sensing images as data sources (Sibanda and Murwira, 2012; Zhong et al., 2014; Peña and Brenning, 2015; Zheng et al., 2015; Waldhoff et al., 2017; Lambert et al., 2018). Since the rise of early agricultural remote sensing technology, these two recognition methods have been widely used. Between them, static identification is usually used for high-resolution aerial imagery and other image data that can only be acquired at restricted times, while dynamic recognition is usually used for more complicated remote sensing classification with temporal image data. The multi-temporal spectral data represent the phenological characteristics of the crop (seasonal dynamics or developmental stages within the year) (Peña and Brenning, 2015). The temporal images can provide useful information about the crop phenology to improve the accuracy of crop classification (Odenweller and Johnson, 1984; Jakubauskas et al., 2002; Knight et al., 2006; Masialeli et al., 2010; Siachalou et al., 2015). However, the success of this method depends largely on the quality and quantity of remote sensing images, and the knowledge level and technical capabilities of decision makers. Studies have shown that it is necessary to use low-dimensional feature space to distinguish land cover types, and that the useful information brought by additional images is usually limited, but the computational complexity and time required are significantly increased (Zhou et al., 2013; Hu et al., 2017). Due to the correlation between features, the increase in the dimension of the data used may reduce

the classification accuracy. This phenomenon is often called the “Hughes effect” (Hughes, 1968; Löw et al., 2013; Pal, 2013). Therefore, it is important to use only a limited number of temporal aerial images to carry out dynamic crop identification. In addition, object-based image analysis (OBIA) methods have developed rapidly and have become an effective method for high-resolution remote sensing image analysis (Hossain and Chen, 2019; Shen et al., 2019). Therefore, users must make choices between pixel-based and object-based classification methods. If the user chooses an object-based classification method, the selection of abundant shapes, textures, and contextual information possessed by image objects needs to be considered.

To solve the above-mentioned problems, this study performed static and dynamic identification to aerial imagery taken from a cropping area using pixel-based and object-based classification. The specific objectives were to: 1) quantify the difference between static identification and dynamic identification in crop classification; 2) evaluate the effect between the object-based and pixel-based classification methods; 3) compare the degrees of influence of layer spectral features, geometry features and texture features of image objects on crop recognition; and 4) identify the best classification methods suitable for crop recognition in aerial remote sensing images under different circumstances.

2 Image acquisition and preprocessing

2.1 Study site

This study was conducted at Texas A&M University’s Research Farm near College Station, Texas, USA (Fig. 1(a)) in June and July 2017. The study area was about 1.61 km² (161 ha) (Fig. 1(b)). The main crop types were cotton, corn, sorghum, and soybean. The area has a humid subtropical climate, with a long frost-free period throughout the year and a longer period of time available for crop growth. Due to different crop varieties, planting conditions, and management methods, there existed great differences in crop growth even for the same crop.

2.2 Image acquisition and preprocessing

2.2.1 Image acquisition

Under good weather conditions, aerial images of the study area were collected on June 16 and July 28, 2017. Two Nikon D7100 digital CMOS cameras with AF Nikkor 24 mm f/2.8D lens (Nikon Inc., Melville, New York) were used to capture images with 6000 × 4000 pixels on a Cessna 206 aircraft. One camera was used to acquire RGB images, and the other was equipped with an 830 nm long-pass filter for acquiring near-infrared (NIR) images. The

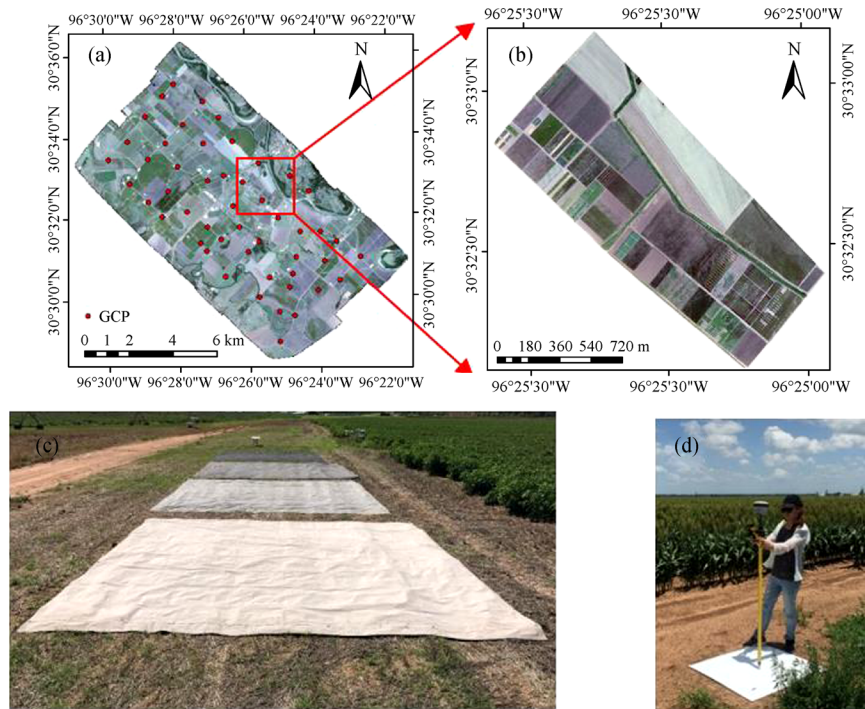


Fig. 1 Location of the study area: (a) airborne RGB mosaic image of July 2017, (b) RGB image of study area, (c) calibration tarps, and (d) ground control point (GCP) measurement with GPS.

flight height was about 1700 m above ground level and the spatial resolution achieved was 0.28 m. The image acquisition was along equidistant flight lines, and the forward overlap and lateral overlap were greater than 70% and 60%, respectively.

2.2.2 Image preprocessing

Pix4D (Pix4D SA, Lausanne, Switzerland) software was used for image mosaicking. To ensure the geographic accuracy of the mosaicked image, 50 white square panels with a side length of 0.6 m were evenly placed in the imaging area as ground control points (GCPs) (Fig. 1(a)), and the coordinates of these GCPs were measured using GPS (Fig. 1(d)). The mosaicked image was converted to the Universal Transverse Mercator (UTM), World Geodetic System (WGS84) geographic coordinate system. Four $8\text{ m} \times 8\text{ m}$ calibration tarpaulins (Fig. 1(c)) with nominal reflectance values of 4%, 16%, 32%, and 48% were placed in the imaging area for radiometric calibration.

The four-band (RGB and NIR) mosaic image in July was selected as the static identification experimental data as crops at this time had more distinct vegetation characteristics. On the basis of this, the June and July normalized difference vegetation index (NDVI) images were combined with the July images to form the six-band dynamic identification experimental data, including the RGB and NIR bands in July, the NDVI in June (6_NDVI), and the NDVI in July (7_NDVI) as shown in Fig. 2.

3 Image classification methods

3.1 Classification process

The classification process was divided into two main parts: pixel-based classification and object-based classification. A machine learning algorithm was introduced and further analysis was carried out in the object features to recognize land cover types under different classification scenarios with higher accuracy. Finally, the best method suitable for crop recognition in aerial remote sensing images was obtained by accuracy assessment. The input images included the two high-resolution aerial images with obvious differences in crop characteristics, and the vegetation characteristics in one of the images (July) were relatively complete. The output was the best classification result under different classification scenarios. The classification flowchart is shown in Fig. 3, and the main steps are described in the following sections.

3.2 Pixel-based classification

The software ENVI was selected for pixel-based classification, and the classification and regression tree (CART) algorithm in machine learning was chosen as the classifier. CART is a decision tree construction algorithm proposed by Breiman et al. (1984). The core idea is to find an optimal feature in the original data set through a number of judgment conditions, gradually dichotomize and refine the

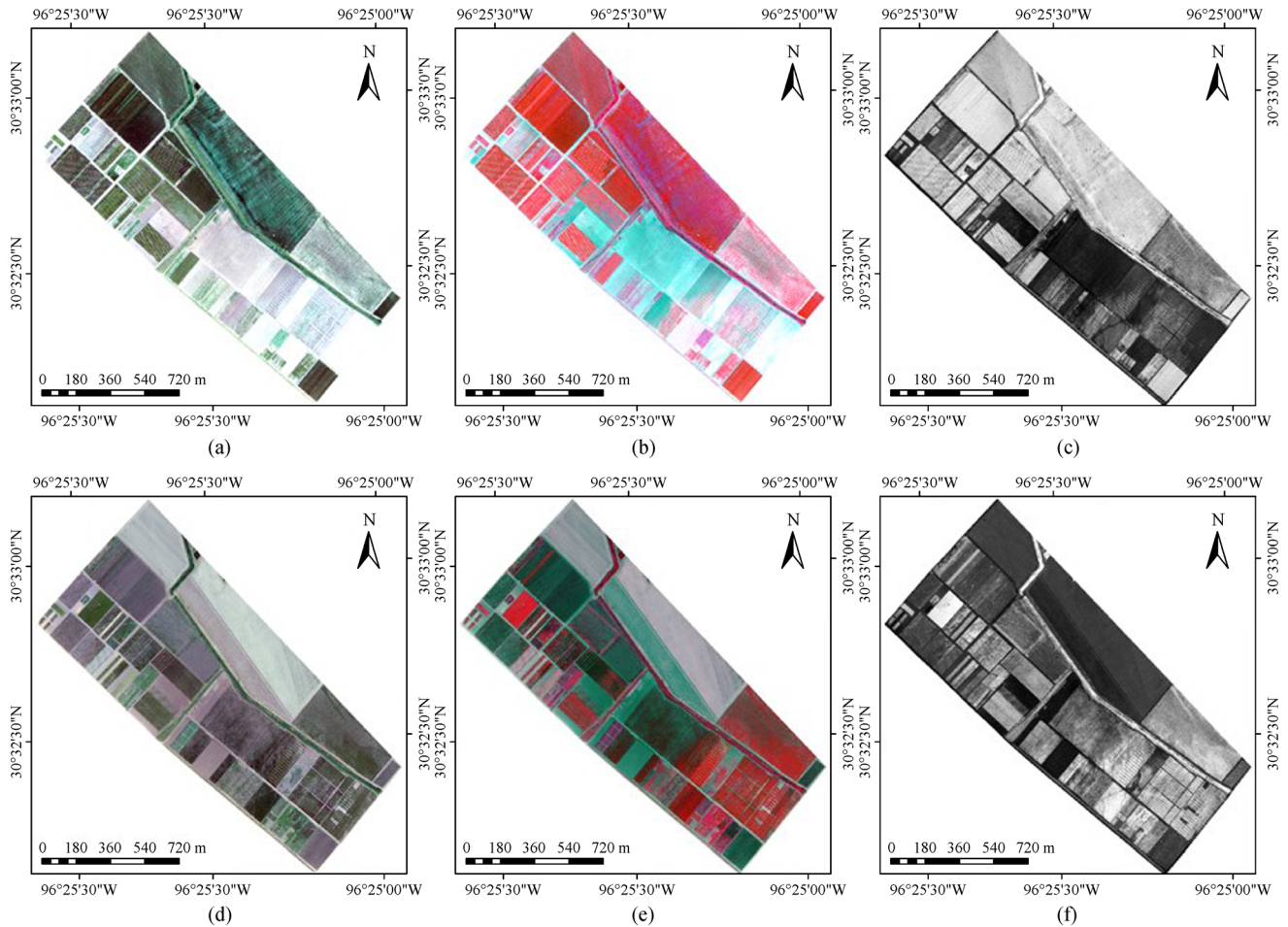


Fig. 2 Airborne images for this study: (a) RGB image of June, (b) Color-infrared (CIR) composite of June, (c) NDVI image of June, (d) RGB image of July, (e) CIR composite of July, and (f) NDVI image of July.

data set, and then repeat the above operations until the conditions for automatic classification of image objects are met. It is a rule-based non-parameter classifier with a “white box” workflow, so the structure and terminal nodes of the decision tree are easy to explain, allowing users to understand the mechanism of the classification method and evaluate it (Zhang et al., 2016).

There were seven general land cover classes in the study area, including bare soil and fallow (BF), corn (CO), cotton (CT), grass (GR), impervious (IM), sorghum (SG), and soybean (SB). Considering the variability within each land cover class, each class was further divided into 3 to 10 subclasses and 5 to 10 training samples were selected from each subclass. The total number of sample samples was about 30 times that of the land cover classes. After supervised classification, these subclasses were merged. The same training samples were used for supervised classification for both the four-band and six-band images.

3.3 Object-based classification

Object-based image analysis has become an effective

method for high-resolution remote sensing image classification (Ma et al., 2017). The commercial software eCognition Developer (Trimble Inc., Munich, Germany) was selected for object-based classification. The process mainly includes two steps: segmentation and classification. The definition and selection of rule sets in each step are the basic and important links in object-based classification (Lichtblau and Oswald, 2019). The CART algorithm was still chosen as the classifier.

3.3.1 Multi-resolution segmentation

Using segmentation algorithms to obtain representative image objects is an important prerequisite for feature extraction, classification, or further integration applications in image analysis (Drăguț et al., 2014). The goal of the multi-scale segmentation algorithm is to minimize the average heterogeneity of the segmented objects (Benz et al., 2004). However, the segmentation scale is an uncertain problem and there is no unique solution (Zhang et al., 2013). To get a more objective segmentation result, the estimation of scale parameter (ESP2) tool (Drăguț

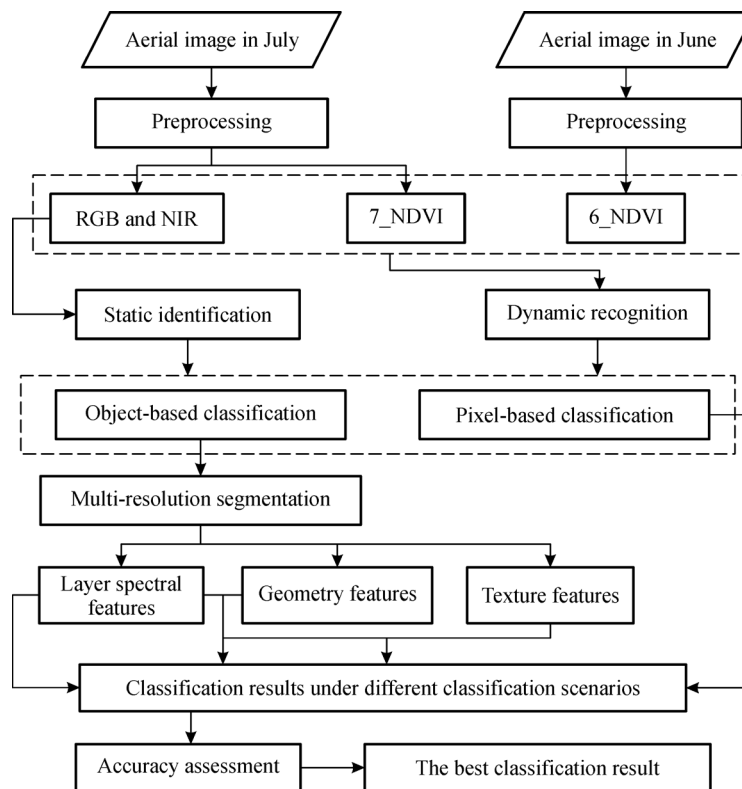


Fig. 3 Flowchart of the classification methods.

et al., 2014) was used to determine the segmentation scale. The ESP2 tool indicates the best segmentation scale by calculating the rate of change of the local variance ($ROC-LV$) of the homogeneity of image objects at different scales (Eq. (1)). When $ROC-LV$ is the largest, that is, when a peak appears, the segmentation scale corresponding to this point is the optimal segmentation scale.

$$ROC-LV = \left(\frac{LV_i - LV_{i-1}}{LV_{i-1}} \right) \times 100, \quad (1)$$

where LV_i represents the mean of the standard deviation (SD) of all image objects in the target image layer and LV_{i-1} represents the mean of standard deviations of all image objects in adjacent low-level image layers.

In this study, a scale step of 10 was set to find the best segmentation scales from 0 to 300 with the ESP2 tool. The best segmentation scales for the four-band and six-band images were 200 and 220, respectively, and a total of 449 (Fig. 4(a)) and 370 (Fig. 4(b)) image objects were generated. To further improve the segmentation quality and solve the problem of over-segmentation of individual features in global image segmentation, a trial-and-error method was used to construct a merged rule set of three typical land classes: bare land, impervious, and vegetation. After merging, the four-band and the six-band images

produced 277 (Fig. 5(a)) and 227 (Fig. 5(b)) image objects, respectively.

3.3.2 Object features

To avoid the impact of different training sample sets on the classification results, the training sample set used for the pixel-based classification was also used in the object-oriented classification. However, the class information contained in the image object was transmitted to the training model. It mainly included layer spectral features (L) including some vegetation indices (VIs), geometry features (G) and texture features (T), as shown in Table 1. The two NDVI layers, 6_NDVI and 7_NDVI, were the unique object characteristics of image objects in dynamic recognition.

However, the number of features and classification accuracy are not simply proportional. Choosing appropriate image features is a key step in any image analysis process (Laliberte et al., 2012). Reducing the dimension of the feature space, that is, selecting a subset of features that has the best effect from a large number of features, can use training samples more effectively and ensure high accuracy of results (Jensen, 2005).

Optimal feature extraction refers to selecting a set of

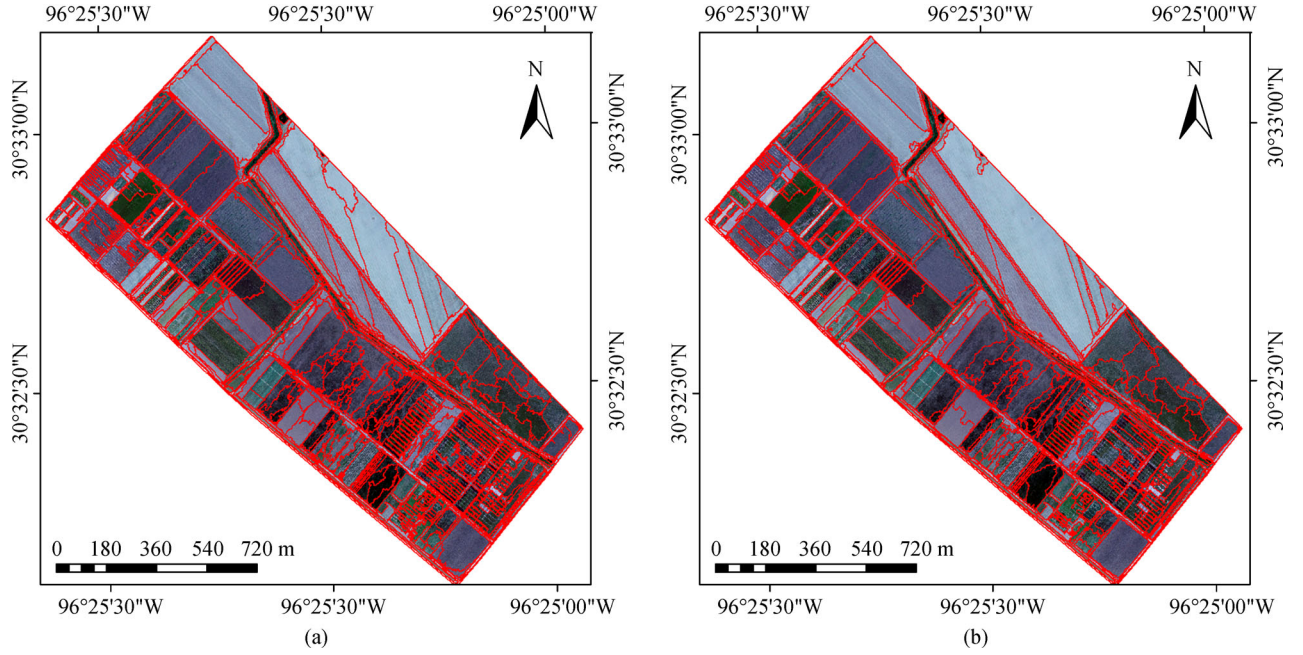


Fig. 4 Global image segmentation results: (a) four-band image segmentation with 449 image objects and (b) six-band image segmentation with 370 image objects.

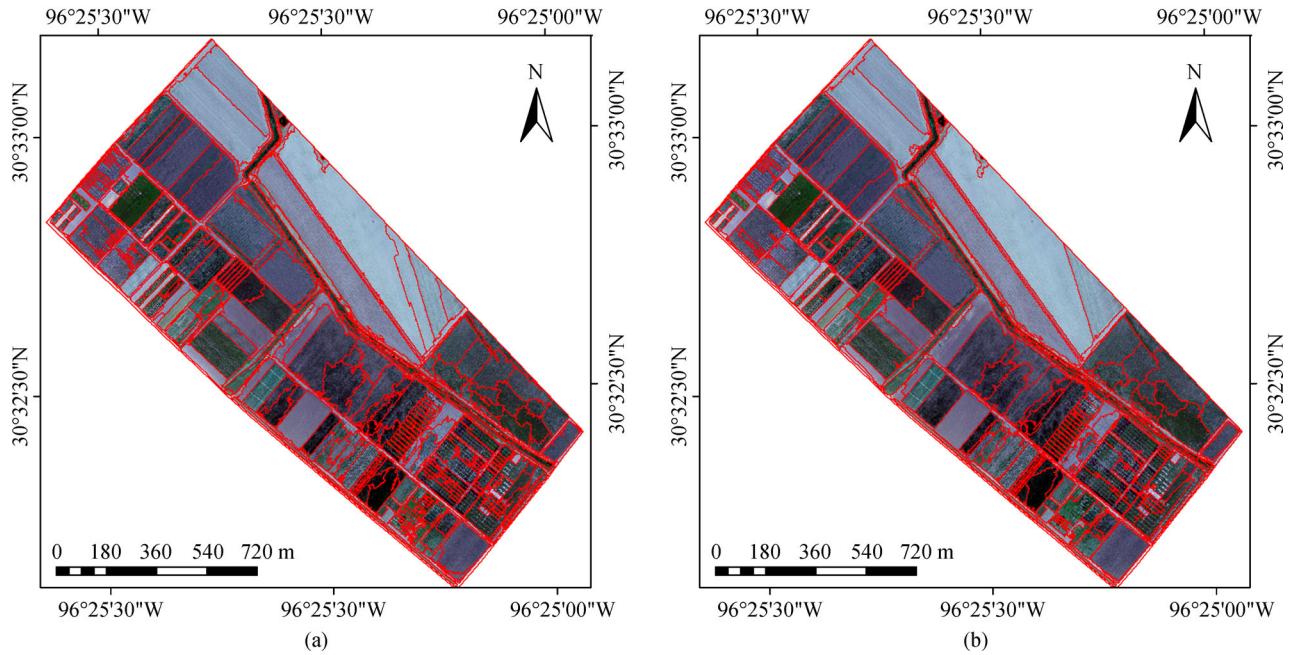


Fig. 5 Image object merged results: (a) four-band image with 277 image objects and (b) six-band image with 277 image objects.

features F with a quantity of $M(N > M)$ from a set of features with a quantity of N to maximize the feature function $D(F)$. The calculation formula is shown in Eq. (2) (eCognition, 2019).

$$D_{ij}(F) = \sqrt{\sum_f \left(\frac{v_f^{(i)} - v_f^{(j)}}{\sigma_f} \right)^2}, \quad (2)$$

where $D_{ij}(F)$ represents the distance between object i and object j under a given feature; $v_f^{(i)}$ represents the feature value of object i to feature f ; $v_f^{(j)}$ represents the feature value of object j to feature f ; σ_f represents the standard deviation of feature f .

If the images are divided into k classes, an k order symmetric matrix D is generated, $D(i,j)$ represents the feature distance $D_{ij}(F)$ corresponding to the two classes.

Table 1 List of object features for classification modeling

source	feature type	feature name	number			
reference	VI	Difference Vegetation Index (DVI) = NIR – R (Richardson and Everitt, 1992)	13			
		Ratio Vegetation Index (RVI) = NI / R (Jordan, 1969)				
		Normalized Difference Vegetation index (NDVI) = (NIR – R) / (NIR + R) (Rouse et al., 1974)				
		D_NDVI = Mean 7_NDVI – Mean 6_NDVI				
		Bare Index (BI) = Blue + Red				
		Soil Adjusted Vegetation Index (SAVI) = 1.5 (NIR – R) / (NIR + R + 0.5) (Huete, 1988)				
		Optimization of Soil-adjusted Vegetation Index (OSAVI) = (NIR – R) / (NIR + R + 0.16) (Rondeaux et al., 1996)				
		B* = B/(B + G + R), G* = G/(B + G + R), R* = R/(B + G + R)				
		Excess Green (EG) = 2G* – R* – B* (Woebbecke et al., 1995)				
		Excess Red (ER) = 1.4R* – G* (Meyer et al., 1999)				
		VI* = EG – ER (Camargo Neto, 2004)				
		eCognition		layer	Mean of B, G, R, NIR, 6_NDVI, 7_NDVI and Brightness \Standard deviation (StdDev) of B, G, R, NIR, 6_NDVI and 7_NDVI HIS (Hue, Intensity, Saturation)	16
				geometry	Area, Border length, Length/Width Asymmetry, Compactness, Density, Shape Index	7
texture	GLMC (Homogeneity, Contrast, Dissimilarity, Entropy, Angular Second Moment, Mean, StdDev, Correlation)		8			
total number of features			44			

$$D = \begin{pmatrix} D_{1,1}(F) & \cdots & D_{1,k}(F) \\ \vdots & \ddots & \vdots \\ D_{k,1}(F) & \cdots & D_{k,k}(F) \end{pmatrix}, \quad (3)$$

where $D_{ij}(F) = D_{ji}(F)$, $D_{ii}(F) = 0$, and the best distance is $MAX(D_{ij}(F))$.

The feature quantities are fully combined as 1,2,...,n, the distance matrix D of each combination is compared, and the optimal feature combination as the feature space for decision tree classification is determined.

Since the value range of D is $[0, \infty)$, it is difficult to accurately measure the separability of a certain feature. Therefore, a variable with a limited value range is introduced, namely the Jeffries-Matusita (J-M) distance (Richards and Jia, 2006). The value range is $[0, 2]$. The calculation formula is shown in Eq. (4).

$$J = 2(1 - e^{-D}), \quad (4)$$

where J is related to the distance matrix D of the feature space. When the J value tends to zero, the feature is less divisible; when the J value approaches 2, the feature is

more divisible. Therefore, according to the magnitude of the J value, the feature space for feature classification can be selected.

3.4 Verification and assessment

To verify the accuracy of the classification results, 500 points were randomly selected in the study area using ArcGIS software, and assigned to each land class based on ground measurement data. Table 2 shows the points and percentages by class type. These points were used to evaluate the accuracy of the classification maps and generate the error matrix of each classification. On this basis, the overall accuracy, Kappa coefficient, producer's accuracy, and user's accuracy were calculated as evaluation indicators (Congalton, 1991; Foody, 2009).

4 Results

4.1 Feature optimization result

For optimal feature extraction, it can be seen from the test

Table 2 Count and percentage by class type for 500 reference points

class type	count	percentage	class type	count	percentage
bare soil and fallow	63	12.6%	impervious	8	1.6%
corn	186	37.2%	sorghum	34	6.8%
cotton	77	15.4%	soybean	63	12.6%
grass	69	13.8%			

results of the sample separability (Fig. 6) that the four-band and six-band image data showed similar changes and trends. With the increase of object features, the separability of samples improved significantly, which is consistent with the expected results. Without considering the classification accuracy, the sample separability was the best for the combination of the layer spectral features, geometry features, and texture features (LGT), followed by the combination of the layer spectral features and geometry features (LG) and the layer spectral features (L). Dimension 7 seemed to be a turning point. When the feature dimension was less than 7, the separability of training samples was not obviously different between the three sets of object features, but the separation distance increased quickly; when the dimension was greater than 7, the separability of the training samples showed distinct differences among the three types of object features, but the separation distance gradually reached a plateau. At high feature dimensions, the separation of the object features was around 1.75 for the four-band image and between 1.75 and 2 for the six-band image. And under the same feature dimension, the separability of the samples in

the six-band image was higher than that in the four-band image. This indirectly verified that the dynamic identification method could improve the ability of crop classification. On the other hand, it can also be seen that the effect of non-spectral features in crop classification was relatively small.

The specific best features used in the subsequent image classification are shown in Table 3. It can be seen from Table 3 that the four-band image used more object features than the six-band image, which could make up for the lack of effective classification information about crop phenology in the single-phase image. In addition to the unique object features of the six-band image, the best feature sets used by the four-band image contained those used by the six-band image in the same feature type. All these illustrated the importance of crop phenology information contained in the six-band image for image classification.

4.2 Classification results

Figure 7 shows the classifications maps of the four-band and six-band images using four different classification

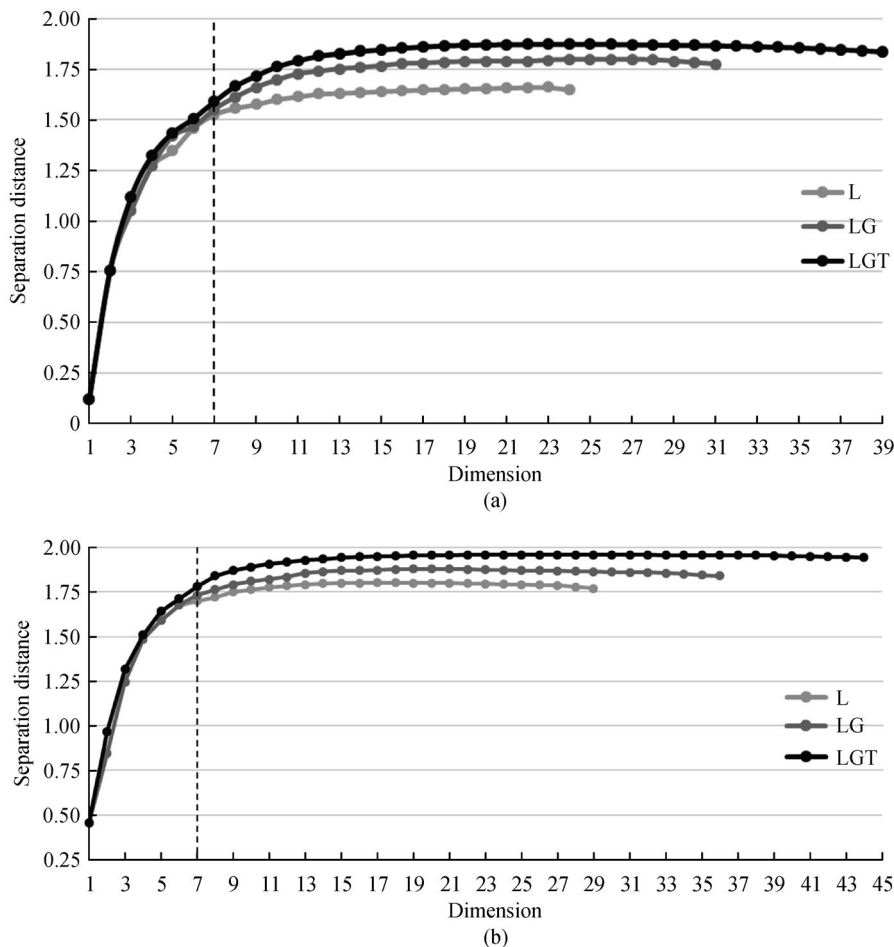


Fig. 6 Feature optimization results of (a) four-band image and (b) six-band image based on three sets of object features: layer spectral features (L), layer spectral features and geometry features (LG), and layer spectral features, geometry features and texture features (LGT).

Table 3 Best object features for different classification scenarios

image	feature type	feature name	number
four-band	L	DVI, RVI, NDVI, BI, SAVI, OSAVI, B*, G*, EG, ER, R*, VI*; Mean of B, G, R, NIR and Brightness; Standard deviation (StdDev) of B, G, R, NIR; HIS (Intensity, Saturation)	23
	LG	DVI, RVI, NDVI, SAVI, OSAVI, B*, G*, EG, ER, R*, VI*; Mean of B, G, R, NIR and Brightness; StdDev of B, G, R, NIR; HIS (Intensity) Area, Border length, Asymmetry, Density, Shape Index	26
	LGT	DVI, RVI, B*, G*, EG, ER, R*, VI*; Mean of NIR; StdDev of B, G, R, NIR; HIS (Intensity) Area, Border length, Asymmetry, Density GLMC (Homogeneity, Dissimilarity, Entropy, Angular Second Moment, StdDev, Correlation)	24
six-band	L	DVI, D_NDVI, B*, G*, EG, ER, R*, VI*; Mean of G, NIR, 6_NDVI and Brightness; StdDev of B, G, NIR, 6_NDVI and 7_NDVI; HIS (Intensity)	18
	LG	DVI, D_NDVI, B*, G*, EG, ER, R*, VI*; Mean of G, NIR, 6_NDVI and Brightness; StdDev of B, G, NIR, 6_NDVI; HIS (Intensity) Area, Asymmetry, Density	20
	LGT	D_NDVI, B*, G*, EG, ER, R*, VI*; Mean of NIR, 6_NDVI; StdDev of B, G, R, NIR, 6_NDVI and 7_NDVI; HIS (Intensity) GLMC (Homogeneity, Contrast, Dissimilarity, Entropy, Angular Second Moment, StdDev, Correlation)	23

Notes: ^{a)} L = the layer spectral features, ^{b)} LG = the layer spectral and geometry features, ^{c)} LGT = the layer spectral, geometry and texture features.

methods. Through visual comparison, all classification maps seemed to be able to effectively distinguish different land cover classes in the area, but the “salt and pepper” effect of the pixel-based classification maps was obvious compared with the object-based classification maps. Due to the phenomena of “different spectra of the same land class” and “different land class of the same spectrum,” the object-based classification maps showed better visual effects. This was also in line with the existing research results (Chubey et al., 2006; Yu et al., 2006; Myint et al., 2011). Without considering the classification accuracy, the object-based classification maps looked cleaner and were more suitable for generating thematic maps. In addition, the visual difference between the four-band and six-band images was also obvious, especially for the pixel-based classifications: the “salt and pepper noise” of the six-band images was alleviated to some extent.

Specifically, the classification result of impervious cover had high consistency and stability. Soybeans were misclassified as corn in the four-band classification maps. Corn had already senesced or was just harvested and some soybean plants were dry and had yellow tone, so both had low chlorophyll content. Corn was also well separated among all the crops due to low chlorophyll. There were some misclassifications between sorghum and cotton. Cotton was in the vegetative stage and sorghum was in the late growth stage. Although the canopy NDVI of sorghum had decreased, the leaves of sorghum were still green and weeds in the background also increased the NDVI of sorghum fields. Similarly, bare soil and fallows were often mistakenly classified as corn in the four-band classification maps. Due to the difference in growth stages and management practices, the four types of crops were confused with grass to some extent. There was also some confusion between cotton and bare soil. In general, all land classes showed different degrees of confusion. However,

this problem did not seem to be as obvious in the object-based classification maps as in the pixel-based supervised classification maps.

4.3 Accuracy assessment

Figure 8 presents the accuracy assessment results of the eight classification maps in intuitive plots. The overall accuracy ranged from 50.4% for 4PB to 88.0% for 6OB-LG, and the overall Kappa coefficient ranged from 0.37 for 4PB to 0.85 for 6OB-LG. As expected, the overall accuracy and Kappa coefficient of the six-band image were higher than those of the four-band image, and the object-based classification was better than pixel-based classification. Relatively speaking, whether it was static recognition or dynamic recognition, the classification effects of the three types of feature sets were small in the object-based classification. The general trends of the mapping accuracy of the four-band and six-band images in the four classification methods could be summarized as OB-L > OB-LG > OB-LGT > PB for the four-band image and OB-LG > OB-LGT > OB-L > PB for the six-band image.

The accuracy assessment results by class for each classification are shown in Table 4. The Kappa coefficients of impervious cover in all classifications were 1, indicating excellent and stable classification for the class. However, the accuracy of grass was relatively low, and the Kappa coefficient in all cases ranged from 0.15 to 0.58 with an average of 0.39. In addition to the 4PB classification, the classification accuracy of bare soil and fallow was relatively high, and the Kappa coefficient was all above 0.72. The accuracy of the corn classification results varied for the four-band image, but the dynamic recognition technology significantly improved its Kappa coefficient to as high as 0.97. The accuracy of soybean in the pixel-based

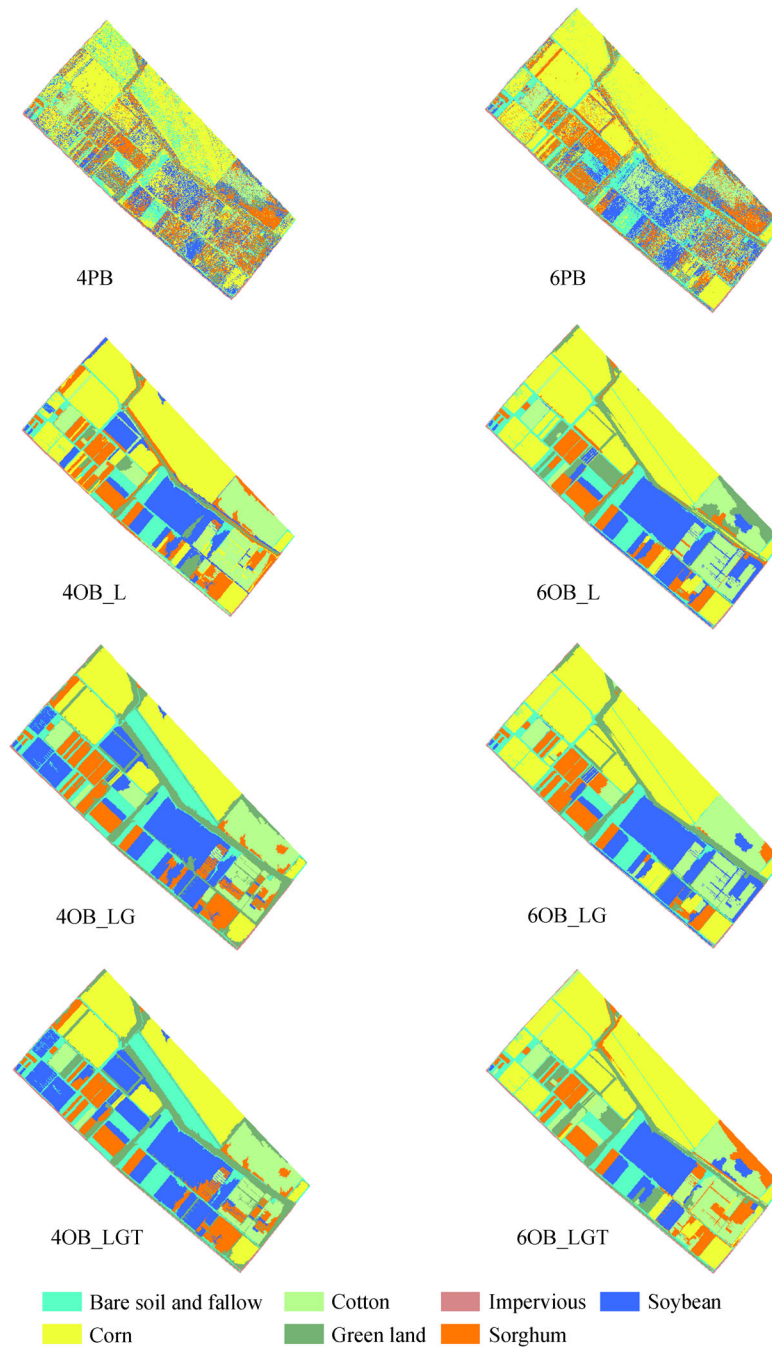


Fig. 7 Image classification results: 4 = four-band image, 6 = six-band image, PB = pixel-based classification, OB-L = object-based classification of the layer spectral features, OB-LG = object-based classification of the layer spectral and geometry features, and OB-LGT = object-based classification of the layer spectral, geometry and texture features.

classifications was low, with a maximum of only 0.43, while the accuracy in the object-based classifications was significantly improved, as high as 0.98. Cotton and sorghum had similar results. The average Kappa coefficient in all classification scenarios was 0.59 for cotton and 0.64 for sorghum. However, if only the object-based classification was considered, the average Kappa coefficients of cotton and sorghum reached 0.71 and 0.75, respectively.

5 Discussion

5.1 Importance of classification methods

To analyze the importance of classification methods on the overall scale, Table 5 presents the average overall accuracy (AOA) and average population Kappa coefficient (AOKp) for both the four-band and six-band image data under the four classification methods. It could be seen that OB-L was

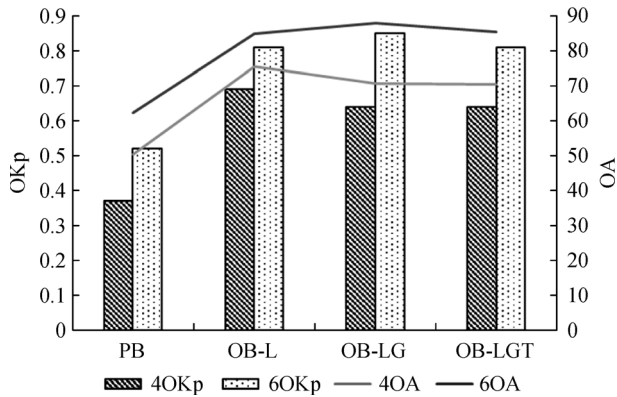


Fig. 8 Results of overall kappa and overall accuracy analysis by four classification methods: OKP = Overall Kappa, OA = Overall accuracy: 4 = four-band image, 6 = six-band image, PB = pixel-based classification, OB_L = object-based classification of the layer spectral features, OB_LG = object-based classification of the layer spectral and geometry features, and OB_LGT = object-based classification of the layer spectral, geometry and texture features.

the best, followed by OB-LG and OB-LGT, and PB was the worst. The difference in classification accuracy among the three feature sets was small. Interestingly, when non-spectral features were added, the classification accuracy was even lower. If the influence of object features on the classification results was not considered, the AOA and AOKp of the object-based classifications were 0.74 and 79.1%, respectively, which were much higher compared with the pixel-based classifications (PB).

From the class scale analysis, Fig. 9 shows the average Kappa coefficient of the two data sources for each class under the four classification methods (AKp1). It could be seen that the impervious class had good classification results under all classification methods, and the classification accuracy of the other classes was significantly improved under the object-based classification method, especially for cotton and soybean. Increasing object features improved the separability of samples, which in turn improved classification accuracy. This conclusion was well reflected in the bare soil and fallow class and the grass

Table 4 Accuracy assessment results by class for four-band and six-band images based on four classification methods

classification		BF	CO	CT	GR	IM	SG	SB
(a) 4PB	Pa (%)	50.8	73.7	29.9	23.2	100	32.4	39.7
	Ua (%)	57.1	69.9	63.9	35.6	80.0	14.7	30.5
	Kp	0.45	0.57	0.24	0.16	1	0.20	0.28
(b) 6PB	Pa (%)	79.4	88.2	32.5	20.3	100	52.9	50.8
	Ua (%)	72.5	86.3	46.3	46.7	88.9	22.2	47.8
	Kp	0.75	0.81	0.24	0.15	1	0.44	0.43
(c) 4OB-L	Pa (%)	74.6	86.6	84.4	24.6	100	82.4	81.0
	Ua (%)	92.2	91.0	76.5	68.0	100	36.8	65.4
	Kp	0.72	0.79	0.81	0.21	1	0.79	0.77
(d) 6OB-L	Pa (%)	95.2	97.8	68.8	53.6	100	94.7	98.4
	Ua (%)	1	94.8	84.1	63.8	100	51.2	81.6
	Kp	0.95	0.97	0.64	0.48	1	0.61	0.98
(e) 4OB-LG	Pa (%)	82.5	60.2	67.5	63.8	100	82.4	90.5
	Ua (%)	61.2	1	82.5	63.8	100	43.8	57.6
	Kp	0.79	0.49	0.63	0.58	1	0.80	0.88
(f) 6OB-LG	Pa (%)	95.2	96.8	89.6	49.3	100	79.4	98.4
	Ua (%)	1	94.7	85.2	89.5	100	57.4	81.6
	Kp	0.95	0.95	0.88	0.45	1	0.77	0.98
(g) 4OB-LGT	Pa (%)	84.1	60.2	67.5	62.3	100	76.5	92.1
	Ua (%)	60.9	1	89.7	62.3	100	46.4	52.7
	Kp	0.81	0.49	0.63	0.56	1	0.74	0.90
(h) 6OB-LGT	Pa (%)	95.2	96.8	72.7	60.9	100	82.4	84.1
	Ua (%)	1	95.7	71.8	79.2	100	51.9	89.8
	Kp	0.95	0.95	0.68	0.56	1	0.80	0.82

Notes: ^{a)} Pa = producer's accuracy, ^{b)} Ua = user's accuracy, ^{c)} Kp = Kappa coefficient, ^{d)} 4 = four-band image, ^{e)} 6 = six-band image, ^{f)} PB = pixel-based classification, ^{g)} OB-L = object-based classification of the layer spectral features, ^{h)} OB-LG = object-based classification of the layer spectral and geometry features, and ⁱ⁾ OB-LGT = object-based classification of the layer spectral, geometry and texture features.

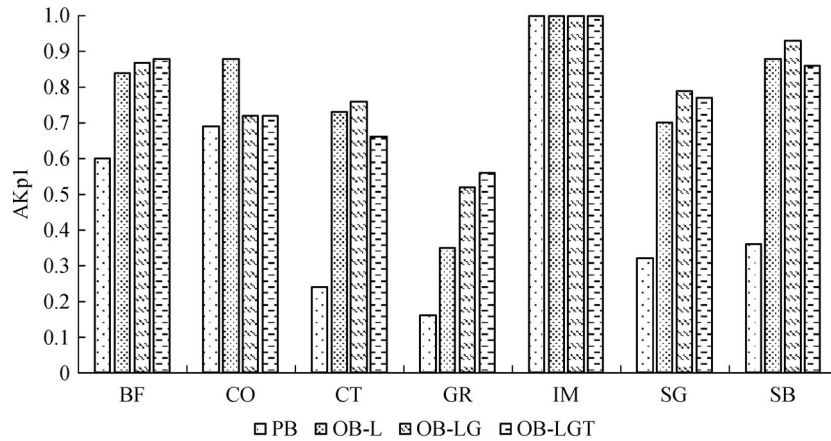


Fig. 9 Analysis of the importance of classification methods by class: BF = bare soil and fallow, CO = corn, CT = cotton, GR = grass, IM = impervious, SG = sorghum, SB = soybean, AKp1 = the average Kappa coefficient of the two data sources for each class under four classification methods, PB = pixel-based classification, OB_L = object-based classification of the layer spectral features, OB_LG = object-based classification of the layer spectral and geometry features, and OB_LGT = object-based classification of the layer spectral, geometry and texture features.

class. For corn, adding non-spectral features reduced classification accuracy, which was because the corn in the area was in the senescing status or just harvested, inclusion of the shape and texture features caused more confusion. For cotton, sorghum, and soybean, inclusion of the texture features lowered the classification accuracy, which was related to the growth of the crops and management conditions. In particular, cotton was planted over a one-month period in the study area, so the appearance of the same crop differed greatly. It also caused the phenomenon that the inclusion of non-spectral features lowered classification accuracy on the overall scale.

5.2 Importance of dynamic recognition

To analyze the importance of dynamic recognition, Fig. 10 shows the average Kappa coefficient of the four classification methods for each class under the two recognition methods (AKp2). From the overall scale, the classification accuracy of the dynamic recognition method was clearly higher than that of the static recognition in all classification methods (Fig. 8). However, on the class scale, the improvement range varied by class (Fig. 10).

Comparing the dynamic recognition with static recognition, the increase of AKp2 for the main crop, corn, was as high as 0.32, while the increase of AKp2 for both cotton

and sorghum was only 0.01. This had much to do with the image acquisition time and crop growth stage. Corn was still lush and green in mid-June, but it started senescing and turning yellow in late June to July. By late July, corn plants became dry and some corn fields were already harvested. The spectral reflectance of the corn fields in the two-date images changed greatly, so it was easy to distinguish using the dynamic recognition method. In contrast, the spectral change in sorghum between the two dates was small, so the dynamic recognition effect was not significant. Because the planting period for cotton spanned over one month, the image characteristics of the early cotton and the late cotton were quite different, which affected the classification results for both recognition methods. The increase of AKp2 for soybean was only 0.07, but its classification accuracy was relatively high. Even under static identification, the average Kappa coefficient for soybean reached 0.73, which was the best among all categories except for impervious cover. Since most of the corn fields were harvested in July, the bare soil and fallow class in the single-date image was easily confused with corn (Fig. 7), which was also one of the main categories for corn misclassification (Table 5). The increase of AKp2 for the bare soil and fallow class was ranked second, indicating that the dynamic identification method could improve the separability of bare soil and fallow.

Table 5 Analysis of the importance of classification methods on the overall scale

	PB	OB-L	OB-LG	OB-LGT
AOKp	0.445	0.750	0.745	0.725
AOA (%)	56.3	80.1	79.3	77.9

Notes: ^{a)} PB = pixel-based classification, ^{b)} OB-L = object-based classification of the layer spectral features, ^{c)} OB-LG = object-based classification of the layer spectral and geometry features, ^{d)} OB-LGT = object-based classification of the layer spectral, geometry and texture features, ^{e)} AOA = the average overall accuracy of the four-band image and six-band image data under the four classification methods, and ^{f)} AOKp = average population Kappa coefficient of the four-band image and six-band image data under the four classification methods.

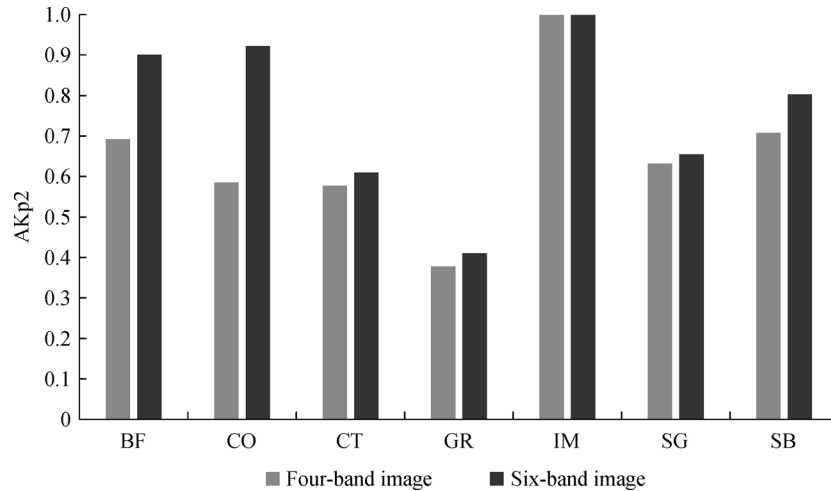


Fig. 10 Analysis of the importance of dynamic recognition by class: BF = bare soil and fallow, CO = corn, CT = cotton, GR = grass, IM = impervious, SG = sorghum, SB = soybean, and AKp2 = the average Kappa coefficient of the four classification methods for each class under the two recognition methods.

5.3 Applicability of classification methods

Through the above analysis, both dynamic recognition and object-based classification could improve the accuracy of aerial image classification for crop recognition. In this study, the same flight platform and imaging system were used to collect image data at two times in the same area. Timing options are advantageous for long-term crop management practitioners, but the operation of the flight platform and the assembly of the imaging system require experienced professionals. The object-based classification method is superior to the pixel-based classification method, and the classification accuracy of the three feature sets differs slightly. Data processing for object-based classification is difficult, and the increase of non-spectral features will make the classification more complex and time-consuming. Moreover, users are required to have a thorough understanding of the classification process. Therefore, cost and user's operation skills are the major limitations of the applicability of the classification method in this study. For experienced and advanced users, a combination of object-based classification considering the non-spectral characteristics and dynamic recognition techniques is a good choice. For the general user, dynamic recognition and object-based classification that consider only spectral features are an appropriate combination. For novice users, dynamic identification techniques can be used for pixel-based classification.

6 Conclusions

Aerial remote sensing for crop identification is a common agricultural remote sensing applications, and this research addressed several important and practical problems related

to it. In this study, machine learning was employed to perform pixel-based and object-based classification on a cropping area with static and dynamic recognition methods. The classification results were further refined based on three types of object features (layer spectral, geometry, and texture). The main conclusions from this study are as follows.

First, static recognition had the highest accuracy of 75.4% in object-oriented classification based on layer spectral features, and dynamic recognition had the highest accuracy of 88.0% in object-oriented classification based on layer spectral and geometry features. Dynamic recognition cannot only reduce the impact of the difference in crop growth and management conditions on the results, but also amplify the differences between different features, thereby improving the accuracy of the classification results.

Second, the object-based classification method is superior to the pixel-based supervised classification method, which is consistent with the expected effect and common sense. The classification accuracy of the three feature sets was relatively small, and the inclusion of non-spectral features could reduce the classification accuracy. It has much to do with external factors such as image acquisition time, crop growth status, and management conditions in the study area.

Finally, when users select image data and classification methods, other limiting factors such as cost, specific application requirements, and operator experience need to be considered. In practical applications, if users do not have advanced image processing capabilities, they can use dynamic recognition for pixel-based classification. If users have higher accuracy requirements and necessary image processing capabilities, they can use a combination of dynamic recognition and object-oriented classification

based on layer spectral and geometry features, or use multiple classification methods for comparative analysis.

In conclusion, the results of this study show that the combination of high spatial resolution aerial images and machine learning-based dynamic recognition techniques can effectively improve crop classification accuracy. The methods and techniques used in this study should provide some practical guidance for crop classification and other agricultural mapping applications. More research is needed to evaluate these methodologies over large cropping areas with more land cover diversity using both aerial and satellite imagery.

Acknowledgements This work was supported by the National Key Research and Development Program (No. 2020YFD1100204), and the Provincial Key Basic Research Project (No. 2019AB002). The authors wish to thank Lee Denham and Fred Gomez at USDA-ARS in College Station, Texas for the collection of the airborne images. Mention of a commercial product is solely for the purpose of providing specific information and should not be construed as a product endorsement by the authors or the institutions with which the authors are affiliated.

References

- Bauer M E, Cipra J E (1973). Identification of agricultural crops by computer processing of ERTS MSS data. In: The Proceedings of the Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. New Carrollton, IN, USA, 3, 205–212
- Benz U C, Hofmann P, Willhauck G, Lingenfelder I, Heynen M (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J Photogramm Remote Sens*, 58(3–4): 239–258
- Boryan C, Yang Z, Mueller R, Craig M (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int*, 26(5): 341–358
- Breiman L, Friedman J, Stone C J, Olshen R (1984). *Classification and Regression Trees*. New York: Wadsworth Inc
- Cai Y, Guan K, Peng J, Wang S, Seifert C, Wardlow B, Li Z (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens Environ*, 210: 35–47
- Camargo Neto J (2004). A combined statistical-soft computing approach for classification and mapping weed species in minimum-tillage systems. Dissertation for the Doctoral Degree. Lincoln: University of Nebraska
- Chubey M S, Franklin S E, Wulder M A (2006). Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogramm Eng Remote Sensing*, 72(4): 383–394
- Congalton R G (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ*, 37(1): 35–46
- Damian J M, Pias O H de C, Cherubin M R, da Fonseca A Z, Fornari E Z, Santi A L (2020). Applying the NDVI from satellite images in delimiting management zones for annual crops. *Sci Agric*, 77(1): e20180055
- Dimov D, Löw F, Uhl J H, Kenjabaev S, Dubovyk O, Ibrakhimov M, Biradar C (2019). Framework for agricultural performance assessment based on MODIS multitemporal data. *J Appl Remote Sens*, 13(2): 1
- Drăguț L, Csillik O, Eisank C, Tiede D (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J Photogramm Remote Sens*, 88(100): 119–127
- eCognition (2019). User Guide. Available at eCognition website
- Farg E, Ramadan M N, Arafat S M (2019). Classification of some strategic crops in Egypt using multi-remotely sensing sensors and time series analysis. *Egypt J Remote Sens Space Sci*, 22(3): 263–270
- Foody G M (2009). Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sens Environ*, 113(8): 1658–1663
- Hossain M D, Chen D (2019). Segmentation for Object-Based Image Analysis (OBIA): a review of algorithms and challenges from remote sensing perspective. *ISPRS J Photogramm Remote Sens*, 150: 115–134
- Hu Q, Wu W B, Song Q, Lu M, Chen D, Yu Q Y, Tang H J (2017). How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? *J Integr Agric*, 16(02): 324–336
- Huete A R (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens Environ*, 25(3): 295–309
- Hughes G F (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans Inf Theory*, 14(1): 55–63
- Jakubauskas M E, Legates D R, Kastens J H (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Comput Electron Agric*, 37(1–3): 127–139
- Jensen J R (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective*. New Jersey: Prentice-Hall, Inc
- Jordan C F (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecological Society of America*, 50(4): 663–666
- Kenduywo B K, Bargiel D, Soergel U (2016). Crop type mapping from a sequence of terrasars-X images with dynamic conditional random fields. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3(7): 59–66
- Knight J F, Lunetta R S, Ediriwickrema J, Khorram S (2006). Regional scale land cover characterization using MODIS-NDVI 250 m multi-temporal imagery: a phenology-based approach. *GIsci Remote Sens*, 43(1): 1–23
- Laliberte A S, Browning D M, Rango A (2012). A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. *Int J Appl Earth Obs Geoinf*, 15: 70–78
- Lambert M J, Traoré P C S, Blaes X, Baret P, Defourny P (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens Environ*, 216: 647–657
- Lichtblau E, Oswald C J (2019). Classification of impervious land-use features using object-based image analysis and data fusion. *Comput Environ Urban Syst*, 75: 103–116
- Löw F, Michel U, Dech S, Conrad C (2013). Impact of feature selection

- on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS J Photogramm Remote Sens*, 85: 102–119
- Ma L, Li M, Ma X, Cheng L, Du P, Liu Y (2017). A review of supervised object-based land-cover image classification. *ISPRS J Photogramm Remote Sens*, 130: 277–293
- Masialeti I, Egbert S, Wardlow B D (2010). A comparative analysis of phenological curves for major crops in Kansas. *GISci Remote Sens*, 47(2): 241–259
- Meyer G E, Hindman T, Laksmi K (1999). Machine vision detection parameters for plant species identification. *Proc SPIE*, 3543: 327–335
- Murmu S, Biswas S (2015). Application of fuzzy logic and neural network in crop classification: a review. *Aquatic Procedia*, 4: 1203–1210
- Myint S W, Gober P, Brazel A, Grossman-Clarke S, Weng Q H (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens Environ*, 115(5): 1145–1161
- Odenweller J B, Johnson K I (1984). Crop identification using Landsat temporal-spectral profiles. *Remote Sens Environ*, 14(1–3): 39–54
- Mutanga O, Dube T, Galal O (2017). Remote sensing of crop health for food security in Africa: potentials and constraints. *Remote Sensing Applications: Society and Environment*, 8: 231–239
- Pal M (2013). Hybrid genetic algorithm for feature selection with hyperspectral data. *Remote Sens Lett*, 4(7): 619–628
- Pal M, Foody G M (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Trans Geosci Remote Sens*, 48(5): 2297–2307
- Peña M A, Brenning A (2015). Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. *Remote Sens Environ*, 171: 234–244
- Richards J A, Jia X (2006). *Remote Sensing Digital Image Analysis*, 3rd ed. Berlin: Springer-Verlag, 273–274.
- Richardson A J, Everitt J H (1992). Using spectral vegetation indices to estimate rangeland productivity. *Geocarto Int*, 7(1): 63–69
- Rouse J W Jr, Haas R H, Schell J A, Deering D W (1974). Monitoring vegetation systems in the great plains with ERTS. In: *Proceedings of the Third ERTS-1 Symposium NASA*, NASA SP-351. Washington: 309–317
- Rondeaux G, Steven M, Baret F (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens Environ*, 55(2): 95–107
- Sakamoto T, Gitelson A A, Nguy-Robertson A L, Arkebauer T J, Wardlow B D, Suyker A E, Verma S B, Shibayama M (2012). An alternative method using digital cameras for continuous monitoring of crop status. *Agric Meteorol*, 154–155: 113–126
- Shen Y, Chen J, Xiao L, Pan D (2019). Optimizing multiscale segmentation with local spectral heterogeneity measure for high resolution remote sensing images. *ISPRS J Photogramm Remote Sens*, 157: 13–25
- Siachalou S, Mallinis G, Tsakiri-Strati M (2015). A hidden markov models approach for crop classification: linking crop phenology to time series of multi-sensor remote sensing data. *Remote Sens*, 7(4): 3633–3650
- Sibanda M, Murwira A (2012). The use of multi-temporal MODIS images with ground data to distinguish cotton from maize and sorghum fields in smallholder agricultural landscapes of Southern Africa. *Int J Remote Sens*, 33(16): 4841–4855
- Song H, Yang C, Zhang J, Hoffmann W C, He D, Thomasson J A (2016). Comparison of mosaicking techniques for airborne images from consumer-grade cameras. *J Appl Remote Sens*, 10(1): 016030
- Torres-Sánchez J, Peña J M, de Castro A I, López-Granados F (2014). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Comput Electron Agric*, 103: 104–113
- van Klompenburg T, Kassahun A, Catal C (2020). Crop yield prediction using machine learning: A systematic literature review. *Comput Electron Agric*, 177: 105709
- van Niel T G, McVicar T R (2004). Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. *Comput Electron Agric*, 45(1–3): 91–108
- Waldhoff G, Lussem U, Bareth G (2017). Multi-Data Approach for remote sensing-based regional crop rotation mapping: a case study for the Rur catchment, Germany. *Int J Appl Earth Obs Geoinf*, 61: 55–69
- Wang P, Fan E, Wang P (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognit Lett*, 141: 61–67
- Woebbecke D M, Meyer G E, Von Bargaen K, Mortensen D A, Woebbecke D M, Meyer G E, von Bargaen K, Mortensen D A (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans ASAE*, 38(1): 259–269
- Wu B, Meng J, Li Q, Yan N, Du X, Zhang M (2014). Remote sensing-based global crop monitoring: experiences with China's CropWatch system. *Int J Digit Earth*, 7(2): 113–137
- Wu M, Yang C, Song X, Hoffmann W C, Huang W, Niu Z, Wang C, Li W, Yu B (2018). Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion. *Sci Rep*, 8(1): 2016
- Yang C, Everitt J H, Murden D (2011). Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Comput Electron Agric*, 75 (2): 347–354
- Yang C, Hoffmann W C (2015). Low-cost single-camera imaging system for aerial applicators. *J Appl Remote Sens*, 9(1): 096064
- Yu Q, Gong P, Clinton N, Biging G, Kelly M, Schirokauer D (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm Eng Remote Sensing*, 72(7): 799–811
- Zhang X, Xiao P, Song X, She J (2013). Boundary-constrained multi-scale segmentation method for remote sensing images. *ISPRS J Photogramm Remote Sens*, 78: 15–25
- Zhang J, Feng L, Yao F (2014). Improved maize cultivated area estimation over a large scale combining MODIS-EVI time series data and crop phenological information. *ISPRS J Photogramm Remote Sens*, 94: 102–113
- Zhang J, Yang C, Song H, Hoffmann W, Zhang D, Zhang G (2016). Evaluation of an airborne remote sensing platform consisting of two consumer-grade cameras for crop identification. *Remote Sens*, 8(3): 257
- Zheng B, Myint S W, Thenkabail P S, Aggarwal R M (2015). A support vector machine to identify irrigated crop types using time-series

- Landsat NDVI data. *Int J Appl Earth Obs Geoinf*, 34: 103–112
- Zhong L, Gong P, Biging G S (2014). Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. *Remote Sens Environ*, 140: 1–13
- Zhou F, Zhang A, Townley-Smith L (2013). A data mining approach for evaluation of optimal time-series of MODIS data for land cover mapping at a regional level. *ISPRS J Photogramm Remote Sens*, 84: 114–129